

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58 ; Search time 269.446 Seconds  
(without alignments)  
7888.487 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_12653\_13951

Perfect score: 1299

Sequence: 1 atgaagacatcttaacc.....taccgcttcgattgttaa 1299

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282.2	98.7	1323	3	US-09-259-214-1
2	1282.2	98.7	1323	3	US-09-318-528-1
3	1282.2	98.7	1323	3	US-09-281-931-1
4	1282.2	98.7	1323	4	US-09-580-515-1
5	1257.4	96.8	1489	4	US-09-540-149A-9
6	1170.2	90.1	1272	2	US-08-910-798-1
7	63	4.8	1266	4	US-09-489-039A-341
8	40.8	3.1	2868	4	US-09-710-794-4
9	36.4	2.8	480	4	US-09-252-991A-15917
10	36.4	2.8	762	4	US-09-252-991A-15888
11	36.4	2.8	885	4	US-09-252-991A-15774
12	36.4	2.7	1068	4	US-09-252-991A-15804
13	35.4	2.7	3076	4	US-09-710-794-1
14	34.6	2.7	2232	4	US-09-902-540-2830
15	34.6	2.7	14570	4	US-09-902-540-1012
16	34.4	2.6	4403765	3	US-09-103-840A-2
17	34.4	2.6	4411529	3	US-09-103-840A-1
18	34.2	2.6	505	4	US-09-621-976-15639
19	34	2.6	998	2	US-07-885-0898-5
20	34	2.6	1044	4	US-09-252-991A-8907
21	34	2.6	1068	4	US-09-710-794-3
22	34	2.6	2313	4	US-09-252-991A-8696
23	34	2.6	2706	4	US-09-252-991A-9082
24	33.8	2.6	12257	4	US-09-949-016-16384
25	33.6	2.6	505	4	US-09-621-976-15639
26	33.4	2.6	399	4	US-09-621-976-8976
27	33	2.5	2511	2	US-08-680-326-116

C 28	33	2.5	35100	2	US-08-770-379-17	Sequence 17, Appl
C 29	33	2.5	35100	3	US-08-757-669A-17	Sequence 17, Appl
C 30	33	2.5	35100	3	US-09-230-377A-17	Sequence 17, Appl
C 31	33	2.5	69701	4	US-09-949-016-14187	Sequence 14187, A
C 32	33	2.5	73308	4	US-09-489-016-16326	Sequence 16326, A
C 33	32.8	2.5	1362	4	US-09-489-039A-2259	Sequence 2259, Ap
C 34	32.8	2.5	1362	4	US-09-489-039A-2538	Sequence 2538, Ap
C 35	32.6	2.5	1584	4	US-09-489-039A-497	Sequence 497, App
C 36	32.4	2.5	84425	4	US-09-949-016-17402	Sequence 17402, A
C 37	32.2	2.5	2430	4	US-09-489-039A-3732	Sequence 3732, Ap
C 38	32.2	2.5	12023	4	US-09-634-238-405	Sequence 405, App
C 39	32.2	2.5	20373	4	US-09-949-016-17086	Sequence 17086, A
C 40	32.2	2.5	20374	4	US-09-949-016-17086	Sequence 17086, A
C 41	32	2.5	1533	4	US-09-489-039A-4607	Sequence 4607, Ap
C 42	32	2.5	11454	4	US-09-949-016-14196	Sequence 14196, A
C 43	32	2.5	11459	4	US-09-949-016-12337	Sequence 12337, A
C 44	31.8	2.4	630	4	US-09-252-991A-6570	Sequence 6570, Ap
C 45	31.8	2.4	1656	4	US-09-252-991A-6620	Sequence 6620, Ap

## ALIGNMENTS

RESULT 1  
US-09-259-214-1  
Sequence 1, Application US/09259214A  
Patent No. 6110719  
GENERAL INFORMATION:  
APPLICANT: Kretz, Keith  
TITLE OF INVENTION: NOVEL PHRYASE  
FILE REFERENCE: DIVER1370-1  
CURRENT APPLICATION NUMBER: US/09/259, 214A  
CURRENT FILING DATE: 1999-03-01  
EARLIER APPLICATION NUMBER: 08/910, 798  
EARLIER FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1323  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1320)  
NAME/KEY: misc\_Feature  
LOCATION: (1)...(1323)  
OTHER INFORMATION: n = A,T,C or G  
US-09-259-214-1

Query Match 98.7% Score 1282.2; DB 3; Length 1323;  
Best Local Similarity 99.3% Pred. No. 0;  
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY	1	ATGAAAGCATCTTAATCCATTTTATCTCTCTGATTCGGTTAACCCGCAATCGCA	60
DB	1	ATGAAAGCATCTTAATCCATTTTATCTCTCTGATTCGGTTAACCCGCAATCGCA	60
QY	61	TTGCGTCAAGTGAAGCCGAGCTGAAGCTGAAAGTGTGATTTGTCAGTCTGATGT	120
DB	61	TTGCGTCAAGTGAAGCCGAGCTGAAGCTGAAAGTGTGATTTGTCAGTCTGATGT	120
QY	121	GTGCGTCTTCAACCAAGGCGACGCACTGATGCAAGATTCACCCGAGCGCATGGCCA	180
DB	121	GTGCGTCTTCAACCAAGGCGACGCACTGATGCAAGATTCACCCGAGCGCATGGCCA	180
QY	181	ACCGGCGGTAACCTGAGTGGTGAACCGCGCGTGTGATGCTATTCGCTATCTC	240
DB	181	ACCGGCGGTAACCTGAGTGGTGAACCGCGCGTGTGATGCTATTCGCTATCTC	240
QY	241	GGAATTACCAAGCCGCGTGTGATGACGCGATTTGTCGGAAGGAGGCTGCGCG	300
DB	241	GGAATTACCAAGCCGCGTGTGATGACGCGATTTGTCGGAAGGAGGCTGCGCG	300

```

QY 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACCCGTTAAACAGGCCAA 360
Db 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACCCGTTAAACAGGCCAA 360
QY 361 GCGCTTCGCGCGGGGCTGGCACCCTGACCTGTGCAATTAACCGTACATTAACAGGAGATAGC 420
Db 361 GCGCTTCGCGCGGGGCTGGCACCCTGACCTGTGCAATTAACCGTACATTAACAGGAGATAGC 420
QY 421 TCCAGTCCCGCATCCGTTATTTAATCTCTTAAAACTGGCGTTTGCCAACTGATTAACGCG 480
Db 421 TCCAGTCCCGCATCCGTTATTTAATCTCTTAAAACTGGCGTTTGCCAACTGATTAACGCG 480
QY 481 AACGTGACTGACGCGCATCTCTGACGAGGGCAGAGAGGTCATTTGCTGACTTTTACCGGGCAT 540
Db 481 AACGTGACTGACGCGCATCTCTGACGAGGGCAGAGAGGTCATTTGCTGACTTTTACCGGGCAT 540
QY 541 CGGCAAAACGGCGTTTGGCGAATGGAACGGGTCCTTAATTTTCCGCAATCAAACTTGTC 600
Db 541 CGGCAAAACGGCGTTTGGCGAATGGAACGGGTCCTTAATTTTCCGCAATCAAACTTGTC 600
QY 601 CTTAAACGTGAGAAACAGAGCAGAAAGCTGTCTTAATTAACGAGGCAATTAACATCGAACTC 660
Db 601 CTTAAACGTGAGAAACAGAGCAGAAAGCTGTCTTAATTAACGAGGCAATTAACATCGAACTC 660
QY 661 AAGGTGAGGCGCGCAAACTGTCTTAAACCGGTGCGTAAAGCTCGCATCAATGCTGAG 720
Db 661 AAGGTGAGGCGCGCAAACTGTCTTAAACCGGTGCGTAAAGCTCGCATCAATGCTGAG 720
QY 721 GAGATATTTCTCTGCAACAGACAGGAAATCCGGAACCGGGTGGGGAAGATCAC 780
Db 721 GAGATATTTCTCTGCAACAGACAGGAAATCCGGAACCGGGTGGGGAAGATCAC 780
QY 781 GATTCACACAGTGAACAACCTTGTAGTTGCAATACGCGCAATTTTATTTGCTACAA 840
Db 781 GATTCACACAGTGAACAACCTTGTAGTTGCAATACGCGCAATTTTATTTGCTACAA 840
QY 841 CGGACCGCCAGAGTTTCCCGCGAGCCCGCAACCCCGTTAATTAATTAATTAAGACAGG 900
Db 841 CGGACCGCCAGAGTTTCCCGCGAGCCCGCAACCCCGTTAATTAATTAATTAAGACAGG 900
QY 901 TTGACCGCCCATCAACCGCAAAACAGGCGTATGCTGACATTAACCACTTCAAGTCTG 960
Db 901 TTGACCGCCCATCAACCGCAAAACAGGCGTATGCTGACATTAACCACTTCAAGTCTG 960
QY 961 TTTATGCGCGGACACGATTAATCTGGCAATCTGGCGGCACTGAGACTCAACTG 1020
Db 961 TTTATGCGCGGACACGATTAATCTGGCAATCTGGCGGCACTGAGACTCAACTG 1020
QY 1021 ACGCTTCCCGGTACGCGGATTAACAGCGCGCAGGTGTAACCTGGGTGTTGAACGCTG 1080
Db 1021 ACGCTTCCCGGTACGCGGATTAACAGCGCGCAGGTGTAACCTGGGTGTTGAACGCTG 1080
QY 1081 CGTCCGCTAAGCATTAACAGCAGTGAATCAAGTTCGTCGTTCTTCCAGACTTTACAG 1140
Db 1081 CGTCCGCTAAGCATTAACAGCAGTGAATCAAGTTCGTCGTTCTTCCAGACTTTACAG 1140
QY 1141 CAGATGCGATTAATAAGCCCGCTGTCTTAATTAACGCGCGCGGAGAGGTTGAACGCTG 1200
Db 1141 CAGATGCGATTAATAAGCCCGCTGTCTTAATTAACGCGCGCGGAGAGGTTGAACGCTG 1200
QY 1201 CTGGCAGGATTAAGAGGGAATGCGCAGGCAATGTTGTTGTTGAGGTTTACGCA 1260
Db 1201 CTGGCAGGATTAAGAGGGAATGCGCAGGCAATGTTGTTGTTGAGGTTTACGCA 1260
QY 1261 ATCGTGAATGAAGCAGCATACCGGCTTGCAAGTTTG 1296
Db 1261 ATCGTGAATGAAGCAGCATACCGGCTTGCAAGTTTG 1296

```

RESULT 2  
 US-09-318-528-1  
 ; Sequence 1, Application US/09318528  
 ; Patent No. 6183740

```

; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHRYASE
; FILE REFERENCE: 09/010/029003
; CURRENT APPLICATION NUMBER: US/09/318,528
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 09/291,931
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214
; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-318-528-1

Query Match      98.7%; Score 1282.2; DB 3; Length 1323;
Best Local Similarity 99.3%; Pred.No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAAGCATCTTAATCCCATTTTATCTCTGATTCCTTAACCCCGCAATCTGCA 60
Db 1 ATGAAAGCATCTTAATCCCATTTTATCTCTGATTCCTTAACCCCGCAATCTGCA 60
QY 61 TTGCTCAGAGTGAACCGGACGCTGAAGCTGGAAGTGTGTGATTTGCTATGCTATGCT 120
Db 61 TTGCTCAGAGTGAACCGGACGCTGAAGCTGGAAGTGTGTGATTTGCTATGCTATGCT 120
QY 121 GTGCGTCTCCCAACCAAGGCCACGCACTGATGACAGATGTCAACCCAGAGCATAGCCA 180
Db 121 GTGCGTCTCCCAACCAAGGCCACGCACTGATGACAGATGTCAACCCAGAGCATAGCCA 180
QY 181 ACCTGCGCGTAAACTGGGTTGCTGACACCGCGCGTGTGATGCTAATGCTTATCTC 240
Db 181 ACCTGCGCGTAAACTGGGTTGCTGACACCGCGCGTGTGATGCTAATGCTTATCTC 240
QY 241 GGAATTAACCAACGCCAGCGTCTGTAAGCCGACGATTCCTGCGCAAAAGGCTGCCG 300
Db 241 GGAATTAACCAACGCCAGCGTCTGTAAGCCGACGATTCCTGCGCAAAAGGCTGCCG 300
QY 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACAGGCGAA 360
Db 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACAGGCGAA 360
QY 361 GCGTTCGCGCGCGGCTGACCTGACCTGTGCAATTAACCGTACATTAACAGGAGATAGC 420
Db 361 GCGTTCGCGCGCGGCTGACCTGACCTGTGCAATTAACCGTACATTAACAGGAGATAGC 420
QY 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCGTTTGCCAACTGATTAACGCG 480
Db 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCGTTTGCCAACTGATTAACGCG 480
QY 481 AACGTGACTGACGCGCATCTCTGACGAGGGCAGAGAGGTCATTTGCTGACTTTTACCGGGCAT 540
Db 481 AACGTGACTGACGCGCATCTCTGACGAGGGCAGAGAGGTCATTTGCTGACTTTTACCGGGCAT 540
QY 541 CGGCAAAACGGCGTTTGGCGAATGGAACGGGTCCTTAATTTTCCGCAATCAAACTTGTC 600
Db 541 CGGCAAAACGGCGTTTGGCGAATGGAACGGGTCCTTAATTTTCCGCAATCAAACTTGTC 600
QY 601 CTTAAACGTGAGAAACAGAGCAGAAAGCTGTCTTAATTAACGAGGCAATTAACATCGAACTC 660
Db 601 CTTAAACGTGAGAAACAGAGCAGAAAGCTGTCTTAATTAACGAGGCAATTAACATCGAACTC 660

```

Qy	661	AAGGTGAGCCGCCGACANTGCTCATTTAAACGGTGTGGTAAGCCCTCGATCAATAGCTGACG	720
Db	661	AAGGTGAGCCGCCGACAAATGCTCATTTAAACGGTGTGGTAAGCCCTCGATCAATAGCTGACG	720
Qy	721	GAGATATTTCTTCTCTGCAACAAGCAACAGGGAATGCCGAGACCCGGGTGTGGGAAGATCAC	780
Db	721	GAGATATTTCTTCTCTGCAACAAGCAACAGGGAATGCCGAGACCCGGGTGTGGGAAGATCAC	780
Qy	781	GATTTCACACCAAGTGGAAACCTTGTGTAAAGTTTGCATTAACGGCGCAATTTTATTTGCTACAA	840
Db	781	GATTTCACACCAAGTGGAAACCTTGTGTAAAGTTTGCATTAACGGCGCAATTTTATTTGCTACAA	840
Qy	841	CGCAGCGCCAGAGGTGGCCCGCAGCGCGGCCACCCCGTTATTAAGATTTGATCAAGACGCG	900
Db	841	CGCAGCGCCAGAGGTGGCCCGCAGCGCGGCCACCCCGTTATTAAGATTTGATCAAGACGCG	900
Qy	901	TTGACGCCCCCATCCACCGCAAAAAACAGCGGTATGTGTGATTAACCACTTCAGTGTG	960
Db	901	TTGACGCCCCCATCCACCGCAAAAAACAGCGGTATGTGTGATTAACCACTTCAGTGTG	960
Qy	961	TTTATCGCCGACACGATATCTAATCTGGCAATCTCGCGCGCGCATCTGGAGCTCAACTGG	1020
Db	961	TTTATCGCCGACACGATATCTAATCTGGCAATCTCGCGCGCGCATCTGGAGCTCAACTGG	1020
Qy	1021	ACGCTTCCCGGTGACGCGGATTAACACGCGCGCAGGTGTGAACGTGTGTTGAACGCTGG	1080
Db	1021	ACGCTTCCCGGTGACGCGGATTAACACGCGCGCAGGTGTGAACGTGTGTTGAACGCTGG	1080
Qy	1081	CGTCGGCTTAACGATTAACAGCCAGTGTGATTCAGGTTTGGCTGTCTTCACAGCTTTTACAG	1140
Db	1081	CGTCGGCTTAACGATTAACAGCCAGTGTGATTCAGGTTTGGCTGTCTTCACAGCTTTTACAG	1140
Qy	1141	CAGATGCGTGTAAACAGCCGCTGTGCATTTAAATACGCCCGCCGAGAGGTAAACTGACC	1200
Db	1141	CAGATGCGTGTAAACAGCCGCTGTGCATTTAAATACGCCCGCCGAGAGGTAAACTGACC	1200
Qy	1201	CTGGCAGAGTGTGAAGAGCGGAATTCGACAGGCGATGTTCGTTGGCAGGTTTAAACGCA	1260
Db	1201	CTGGCAGAGTGTGAAGAGCGGAATTCGACAGGCGATGTTCGTTGGCAGGTTTAAACGCA	1260
Qy	1261	ATCGTAATGAAGACACGACATACCCGCTTGACGTTTG	1296
Db	1261	ATCGTAATGAAGACACGACATACCCGCTTGACGTTTG	1296

Query Match		98.7%;	Score 1282.2;	DB 3;	Length 1323;
Best Local Similarity		99.3%;	Prod. No. 0;		
Matches 1287;		Conservative	0;	Mismatches	9;
				Indels	0;
				Gaps	0;
Qy	1	ATGAAAGCCATCTTAAATCCCATTTTATCTTTCTGATTCGGTTAACCCCGCAATCTGCA	60		
Db	1	ATGAAAGCGATCTTAAATCCCATTTTATCTTCTGTATTCGGTTAACCCCGCAATCTGCA	60		
Qy	61	TTGGCTCAGATGATGACCGGAGCTGAAAGCTGGAAAGTGTGTGATTGTCAATCGTCAGT	120		
Db	61	TTGGCTCAGATGATGACCGGAGCTGAAAGCTGGAAAGTGTGTGATTGTCAATCGTCAGT	120		
Qy	121	GTGCGTCTCCAAACCAAGGCGACGCAACGTATGCGAGATGTCAACCCGACGATGGCCA	180		
Db	121	GTGCGTCTCCAAACCAAGGCGACGCAACGTATGCGAGATGTCAACCCGACGATGGCCA	180		
Qy	181	ACCTGGCCGGTAAATCTGGGTTGGCTGACACCGCGGGTGGAGCTAATCGCTTATCTC	240		
Db	181	ACCTGGCCGGTAAATCTGGGTTGGCTGACACCGCGGGTGGAGCTAATCGCTTATCTC	240		
Qy	241	GGAATTAAACCAAGCCAGCGCTGTGTAGCCGACGGATTTGCTGGAAAAAGGCTGCCG	300		
Db	241	GGAATTAAACCAAGCCAGCGCTGTGTAGCCGACGGATTTGCTGGAAAAAGGCTGCCG	300		
Qy	301	CAGTCTGTGATGAGTCCGATTTATTTGCTGATGTGACGACGGGTACCCCTAAACAGGGGAA	360		
Db	301	CAGTCTGTGATGAGTCCGATTTATTTGCTGATGTGACGACGGGTACCCCTAAACAGGGGAA	360		
Qy	361	GCCTTGGCGCGCGGGGCTGGACCGTACGTGTGCAATTAACGGTACATCCAGGCGATATCG	420		
Db	361	GCCTTGGCGCGCGGGGCTGGACCGTACGTGTGCAATTAACGGTACATCCAGGCGATATCG	420		
Qy	421	TCCAGTCCCGATCCGTTATTTAATCTCTAATACTGGCGTTTGGCAATGATTAACGG	480		
Db	421	TCCAGTCCCGATCCGTTATTTAATCTCTAATACTGGCGTTTGGCAATGATTAACGG	480		
Qy	481	AACGTGACTGACCGGATCTCAGCAGGGGACAGAGGGTCAATTGCTGACTTACCGGGCAT	540		
Db	481	AACGTGACTGACCGGATCTCAGCAGGGGACAGAGGGTCAATTGCTGACTTACCGGGCAT	540		
Qy	541	CGGCAACCGCGTTTCCGCGAATCGGAAACGGGTGCTTAATTTCCGCAATCAACTGTGC	600		
Db	541	CGGCAACCGCGTTTCCGCGAATCGGAAACGGGTGCTTAATTTCCGCAATCAACTGTGC	600		
Qy	601	CTTAAAGCTGAGAAACAGACAGAAAGCTGTTCATTAAACGACGATTAACATCGAACTC	660		
Db	601	CTTAAAGCTGAGAAACAGACAGAAAGCTGTTCATTAAACGACGATTAACATCGAACTC	660		
Qy	661	AAGGTGACGCGCGCAATATCTCTAATTAACGGGTGCGTAAAGCTTCGATCAATCTGACG	720		
Db	661	AAGGTGACGCGCGCAATATCTCTAATTAACGGGTGCGTAAAGCTTCGATCAATCTGACG	720		
Qy	721	GAGATATTTCTCTGCAACAAACAGAGGAAATGCCGAGCCGGGGTGGGAAAGATCAAC	780		
Db	721	GAGATATTTCTCTGCAACAAACAGAGGAAATGCCGAGCCGGGGTGGGAAAGATCAAC	780		
Qy	781	GATTCACACCAAGTGAACACCTGTCTAAGTTGTGATTAACGCGCAATTTATTTGCTACAA	840		
Db	781	GATTCACACCAAGTGAACACCTGTCTAAGTTGTGATTAACGCGCAATTTATTTGCTACAA	840		
Qy	841	CGACAGCGCAGAGTGTGCGGACGCGCGCACCCCGTTATTTAGATTGATCAACAGCG	900		
Db	841	CGACAGCGCAGAGTGTGCGGACGCGCGCACCCCGTTATTTAGATTGATCAACAGCG	900		
Qy	901	TTGACGCGCCATCCACCGCAAAAACAGCGTATGTGTGATTAACCATTCACGTGCTG	960		
Db	901	TTGACGCGCCATCCACCGCAAAAACAGCGTATGTGTGATTAACCATTCACGTGCTG	960		
Qy	961	TTTATTCGCGGACACAGATCTAATCTGTGGCAAAATCTGGCGCGGCACTGGAGCTCAACTGG	1020		
Db	961	TTTATTCGCGGACACAGATCTAATCTGTGGCAAAATCTGGCGCGGCACTGGAGCTCAACTGG	1020		

QY 1021 ACGTTTCCCGGTACCGGATTAACACGCCGCCAGGTGGTGAACGTGTTTGAACGCTGG 1080  
Db 1021 ACGTTTCCCGGTACCGGATTAACACGCCGCCAGGTGGTGAACGTGTTTGAACGCTGG 1080  
QY 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGCTGCTTCACAGCTTTACG 1140  
Db 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGCTGCTTCACAGCTTTACG 1140  
QY 1141 CAGATGCGTGAATAACCGCGCTGCTATTAAATACGCCGCCGAGAGGTGAACCTGAC 1200  
Db 1141 CAGATGCGTGAATAACCGCGCTGCTATTAAATACGCCGCCGAGAGGTGAACCTGAC 1200  
QY 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGCATGTGTTGTTGCGAGGTTTACGCA 1260  
Db 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGCATGTGTTGTTGCGAGGTTTACGCA 1260  
QY 1261 ATCGTAATGAAGACGACATACCGGCTTGCACTTTG 1296  
Db 1261 ATCGTAATGAAGACGACATACCGGCTTGCACTTTG 1296

RESULT 4  
US-09-580-515-1  
; Sequence 1, Application US/09580515  
; Patent No. 6720014  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/09/580,515  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; PRIOR APPLICATION NUMBER: 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1320)  
; NAME/KEY: misc feature  
; LOCATION: (1)...(1323)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-580-515-1

Query Match 98.7%; Score 1282.2; DB 4; Length 1323;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGTATTCGTTAACCCCGCAATCTGA 60  
Db 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGTATTCGTTAACCCCGCAATCTGA 60  
QY 61 TTCCGCTCAGAGTGAAGCGGAGCTGAAGCTGGAAGTGTGTGATTTGTCACTGCTATGGT 120  
Db 61 TTCCGCTCAGAGTGAAGCGGAGCTGAAGCTGGAAGTGTGTGATTTGTCACTGCTATGGT 120  
QY 121 GTGCGTGTCTCAACCAAGGCAAGCAACTGATGCAAGATGTCAACCCAGAGCGATGGCA 180  
Db 121 GTGCGTGTCTCAACCAAGGCAAGCAACTGATGCAAGATGTCAACCCAGAGCGATGGCA 180  
QY 181 ACCTGGCCGGTAAACTGGGTTGGCTGACACCGCGCGGTGTGAGCTTAATCGCTATCTC 240  
Db 181 ACCTGGCCGGTAAACTGGGTTGGCTGACACCGCGCGGTGTGAGCTTAATCGCTATCTC 240

QY 241 GAACTTACCAAGCCAGCGCTCTGTAGCCGACGGAATGTGCGGAAAAAGGCTGCCG 300  
Db 241 GAACTTACCAAGCCAGCGCTCTGTAGCCGACGGAATGTGCGGAAAAAGGCTGCCG 300  
QY 301 CAGTGTGTCAGGTGCGGATTTATGCTGATGTCAAGAGCTTACCCGTAACAGGCGAA 360  
Db 301 CAGTGTGTCAGGTGCGGATTTATGCTGATGTCAAGAGCTTACCCGTAACAGGCGAA 360  
QY 361 GCGTTCCGCGCGCGGCTGGGCACTGACGTGTCAATTAACCGTAATCCAGGAGATACG 420  
Db 361 GCGTTCCGCGCGCGGCTGGGCACTGACGTGTCAATTAACCGTAATCCAGGAGATACG 420  
QY 421 TCAGTCCCGATTCGTTATTTAATCTCTTAATAAACTGCGTTTGCACTGATTAACGCG 480  
Db 421 TCAGTCCCGATTCGTTATTTAATCTCTTAATAAACTGCGTTTGCACTGATTAACGCG 480  
QY 481 AACGTACTGAAGCGATCTCAGACGAGGAGGTCATTTGCTGACTTTACCGGGCAT 540  
Db 481 AACGTACTGAAGCGATCTCAGACGAGGAGGTCATTTGCTGACTTTACCGGGCAT 540  
QY 541 CGGCAAAAGCGGTTTCCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAAACCTGTC 600  
Db 541 CGGCAAAAGCGGTTTCCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAAACCTGTC 600  
QY 601 CTTAAACGTGAAGAAACAGAGAAAGCTGTTCAATTAACGAGGCAATTCAGATCGGAATC 660  
Db 601 CTTAAACGTGAAGAAACAGAGAAAGCTGTTCAATTAACGAGGCAATTCAGATCGGAATC 660  
QY 661 AAGGTAGCGCGGACATGTCTCAATTAACCGGTGCGTGAAGCTTCGATCAATGCTGACG 720  
Db 661 AAGGTAGCGCGGACATGTCTCAATTAACCGGTGCGTGAAGCTTCGATCAATGCTGACG 720  
QY 721 GAGATATTTCTCCGCAACAGCAAGGAAATGCGGAGCGGGGTGGGAAAGATCAC 780  
Db 721 GAGATATTTCTCCGCAACAGCAAGGAAATGCGGAGCGGGGTGGGAAAGATCAC 780  
QY 781 GATTCAACACAGTGAACACCTTGCTAAGTTGCAATACGCGCAATTTTATTTGTACAA 840  
Db 781 GATTCAACACAGTGAACACCTTGCTAAGTTGCTAAGTTGCAATACGCGCAATTTTATTTGTACAA 840  
QY 841 CGCAGCGCAAGGTTGCCCGCAGCGCCGACCCCGTATTTGATTTGATCAAGCAGCG 900  
Db 841 CGCAGCGCAAGGTTGCCCGCAGCGCCGACCCCGTATTTGATTTGATTTGATCAAGCAGCG 900  
QY 901 TTGAGCGCCCATCCACCGCAAAAACAGGCGTATGTGATGACATTAACCACTCAGTCTG 960  
Db 901 TTGAGCGCCCATCCACCGCAAAAACAGGCGTATGTGATGACATTAACCACTCAGTCTG 960  
QY 961 TTTATGCGCGGACAGATACATATCTGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020  
Db 961 TTTATGCGCGGACAGATACATATCTGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020  
QY 1021 ACGTTTCCCGGTACCGCGGATTAACAGCGCGCGAGGTGTGAACGTGTTGTAAGCTGG 1080  
Db 1021 ACGTTTCCCGGTACCGCGGATTAACAGCGCGCGAGGTGTGAACGTGTTGTAAGCTGG 1080  
QY 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGCTGCTTCACAGCTTTACG 1140  
Db 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGCTGCTTCACAGCTTTACG 1140  
QY 1141 CAGATGCGTGAATAACCGCGCTGCTATTAAATACGCCGCCGAGAGGTGAACCTGAC 1200  
Db 1141 CAGATGCGTGAATAACCGCGCTGCTATTAAATACGCCGCCGAGAGGTGAACCTGAC 1200  
QY 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGCATGTGTTGTTGCGAGGTTTACGCA 1260  
Db 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGCATGTGTTGTTGCGAGGTTTACGCA 1260  
QY 1261 ATCGTAATGAAGACGACATACCGGCTTGCACTTTG 1296  
Db 1261 ATCGTAATGAAGACGACATACCGGCTTGCACTTTG 1296



RESULT 5  
US-09-540-149A-9  
Sequence 9, Application US/09540149A  
Patent No. 6511699  
GENERAL INFORMATION:  
APPLICANT: lei, Xingen  
TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY  
FILE REFERENCE: 19603/2791  
CURRENT APPLICATION NUMBER: US/09/540,149A  
CURRENT FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: 60/127,032  
PRIOR FILING DATE: 1999-03-31  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 9  
LENGTH: 1489  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-540-149A-9

Query Match 96.8%; Score 1257.4; DB 4; Length 1489;  
Best Local Similarity 98.0%; Pred. No. 0;  
Matches 1273; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 ATGAAAGCCATCTTATCCCATTTTATCTCTCTGATTCGGTTAACCCCGCATCTGCA 60  
DB 182 ATGAAAGCCATCTTATCCCATTTTATCTCTCTGATTCGGTTAACCCCGCATCTGCA 241  
QY 61 TTGGCTCAGATGAGCGCGAGCTGAAGCTGAAAAGTGGTGAATTGTCAGTCTGATGT 120  
DB 242 TTGGCTCAGATGAGCGCGAGCTGAAGCTGAAAAGTGGTGAATTGTCAGTCTGATGT 301  
QY 121 GTGCGTCTCAACCAAGCCAGCACTGATGAGGATGTCAACCCGAGCGATGCGCA 180  
DB 302 GTGCGTCTCAACCAAGCCAGCACTGATGAGGATGTCAACCCGAGCGATGCGCA 361  
QY 181 ACCTGGCCGCTAAAACCTGGTGGCTGACACCGCGCGTGTGAGCTAATCGCTATCTC 240  
DB 362 ACCTGGCCGCTAAAACCTGGTGGCTGACACCGCGCGTGTGAGCTAATCGCTATCTC 421  
QY 241 GGAATTACCAAGCCAGCGCTGTGAGCGGATGTCGGGGAAGGCGCTGCCG 300  
DB 422 GGAATTACCAAGCCAGCGCTGTGAGCGGATGTCGGGGAAGGCGCTGCCG 481  
QY 301 CAGTCTGTCAGGTGCGGATTAATGCTGATGTCGACGAGGTAACCCGTAAACAGCGGAA 360  
DB 482 CAGTCTGTCAGGTGCGGATTAATGCTGATGTCGACGAGGTAACCCGTAAACAGCGGAA 541  
QY 361 GCCTTCCGCGCGCGGCTGCGACCTGACTGTGCAATACCGTAATACCCAGGCGATACG 420  
DB 542 GCCTTCCGCGCGCGGCTGCGACCTGACTGTGCAATACCGTAATACCCAGGCGATACG 601  
QY 421 TTCAGTCCGATCCGCTATTATTAACCTCTAAACCTGGCGTTTCCCAATGATGATGCGG 480  
DB 602 TTCAGTCCGATCCGCTATTATTAACCTCTAAACCTGGCGTTTCCCAATGATGATGCGG 661  
QY 481 AACGTGATGACGCGATCTCAGACGAGGACAGAGGTCATATGCTGAATTTACCGGCAT 540  
DB 662 AACGTGATGACGCGATCTCAGACGAGGACAGAGGTCATATGCTGAATTTACCGGCAT 721  
QY 541 CGGCAAAACGCGGTTTCCGGAACCTGAAACGCGGCTTAAATTTTCCGCAATCAAATTTGTC 600  
DB 722 CGGCAAAACGCGGTTTCCGGAACCTGAAACGCGGCTTAAATTTTCCGCAATCAAATTTGTC 781  
QY 601 CTTAAAGCTGAGAAACGAGAAAGCTGATTAACGAGGATTAACATCGGAATC 660  
DB 782 CTTAAAGCTGAGAAACGAGAAAGCTGATTAACGAGGATTAACATCGGAATC 841  
QY 661 AAGGTGAGCGCGCAATGCTCATTAACCGGTGCGGTAACTCTGCAATCATGCTGACG 720  
DB 842 AAGGTGAGCGCGCAATGCTCATTAACCGGTGCGGTAACTCTGCAATCATGCTGACG 901  
QY 721 GAGATATTCTCTCGCAACAGCAAGGAAATCCCGAGCGCGGCTGGGGAAGATCAC 780

DB 902 GAAATATTCTCTCTGCAACAGCAAGGAAATCCCGAGCCGCGGTGGGAAAGATCACT 961  
QY 781 GATTCACACCGTGAACACCTTGTCTAATGTTGATTAAGCCGCAATTTATTTGCTACAA 840  
DB 962 GATTACACCAAGTGAACACCTTGTCTAATGTTGATTAAGCCGCAATTTATTTGCTACAA 1021  
QY 841 CGCAGCGCAGAGTTGCCCGCGAGCGCGCACCCCGTATTAGATTGATTAAGACAGCG 900  
DB 1022 CGCAGCGCAGAGTTGCCCGCGAGCGCGCACCCCGTATTAGATTGATTAAGACAGCG 1081  
QY 901 TTGACGCCCATTCACCGCAAAACAGCGGTATGTTGATTAACATTAACCACTTCAGTCTG 960  
DB 1082 TTGACGCCCATTCACCGCAAAACAGCGGTATGTTGATTAACATTAACCACTTCAGTCTG 1141  
QY 961 TTTATGCGCGACACATTAATCTTGGCAATCTGGCGGCGCATCTGAGTCAACTG 1020  
DB 1142 TTTATGCGCGACACATTAATCTTGGCAATCTGGCGGCGCATCTGAGTCAACTG 1201  
QY 1021 ACCCTTCCCGGTCAAGCGGATTAACAGCGCGCAAGTGAATCTGGTGTGAACGCTGG 1080  
DB 1202 ACCCTTCCCGGTCAAGCGGATTAACAGCGCGCAAGTGAATCTGGTGTGAACGCTGG 1261  
QY 1081 CGTGGCTAAGCGATTAACAGCGGATTAAGTTCAGTGTCTTCCAGACTTTACAG 1140  
DB 1262 CGTGGCTAAGCGATTAACAGCGGATTAAGTTCAGTGTCTTCCAGACTTTACAG 1321  
QY 1141 CAGATCGGTATTAACCGCGCTGTCAATTAATACCGCGCGCAAGTGAATCTGAC 1200  
DB 1322 CAGATCGGTATTAACCGCGCTGTCAATTAATACCGCGCGCAAGTGAATCTGAC 1381  
QY 1201 CTGCGAGATGTAAGCAAGCAATGCGAGGCGATGTTGTTGCGAGGTTTACGCA 1260  
DB 1382 CTGCGAGATGTAAGCAAGCAATGCGAGGCGATGTTGTTGCGAGGTTTACGCA 1441  
QY 1261 ATCGTAATGAACAGCATACCGGCTTGGCAATTTGTA 1299  
DB 1442 ATCGTAATGAACAGCATACCGGCTTGGCAATTTGTA 1480

RESULT 6  
US-08-910-798-1  
Sequence 1, Application US/08910798  
Patent No. 5876997  
GENERAL INFORMATION:  
APPLICANT: KREIZ  
TITLE OF INVENTION: NOVEL PHYTASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,798  
FILING DATE: August 13, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, PH.D., LISA A.  
REGISTRATION NUMBER: 38,347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5099  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

Query Match	Best Local Similarity	Score	DB #	Length
Matches 202;	Conservative	4.8%;	48.7%;	25-10;
				Pred. No. 26;
				Mismatches 210; Indels 3; Gaps 1
QY	86	AGCTGGAAGATGTGTGATATTGTCAGTCTTCATAGTGTGTGGTCTCC---AACCAAGGCCA	142	
Db	113	AGCTGCGACGAGGTGCTGATCATAGGCGGTGCACACTGTGTGCGCGTGGCCATTAACG	172	
QY	143	CGCAACTGATGCAGATGTATCATCCCGACGCATATGGCCAACTGGCCGGTAAATCTGGTT	202	
Db	173	GCAGCGTGTGGAAACAGTCCACCGCCAGGCGCTGGGCGCGAGTGGATGTGCCGGGGGCTC	232	
QY	203	GGCTGACACCGCCCGGTGTGTGAATTAATGGCCATATCGGACATTTCCAAAGCCAGAGCTC	262	
Db	233	AGCTGACCAACAAAGGGGTGTCTGAGGTGTATATGGGCCACTTATATGCCGATATGGC	292	
QY	263	TGGTAGCCGACGAGATTGCTGCGCAAAAAGGCGTCCCGCAGTCTGTCAAGTTCGCATTAT	322	
Db	293	TGGGCGACAGAAAGCTGTGTATGCCAGCGGCGAGTGCCTCGCGGAAAAACGCGTTATATGCTT	352	
QY	323	TTGCTGATGTTCGACGAGAGCTTACCCCGTAAAAACAGCGAAGCTTCGCGCGCGGCTGGCAC	382	

Db 353 ACCGTTACACGCTCCGACGCCGACCGCTGCGCACCGCGAGTCTTTATCATCCGCGCGCTTC 412  
Qy 383 CTGACTGTGCATTAACCGTACATACCGAGCAGATACGTCCAGTCCGATTCTTATTA 442  
Db 413 CCGGCGTGGGACATCCGGTGCATCACAGCTCAGATGGGACATGACACCGACCTTCA 472  
Qy 443 ATTCCTTAATACTGGCGTTTGGCAACTGATTAACCGCAACGTGACTGACGCGAT 497  
Db 473 ACCCGGTTATTAACGACGACTCCCGCGCTTTCGCGAAGGCGCTGACGCGAT 527

## RESULT 8

US-09-710-794-4/c  
; Sequence 4, Application US/09710794  
; Patent No. 6573069  
; GENERAL INFORMATION:  
; APPLICANT: Holloway, James L.  
; APPLICANT: Gao, Zeren  
; APPLICANT: Whitmore, Theodore E.  
; TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMS1  
; FILE REFERENCE: 99-76  
; CURRENT APPLICATION NUMBER: US/09/710,794  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/164,685  
; PRIOR FILING DATE: 1999-11-10  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 2868  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (173)...(1219)  
US-09-710-794-4

Query Match 3.1%; Score 40.8; DB 4; Length 2868;  
Best Local Similarity 53.8%; Pred. No. 0.012;  
Matches 84; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 128 CTCGACCAAGGCGACGCACTGATGAGATGCACCCGACGATGGCCAACTGGC 187  
Db 353 CTCCTCTCCCTGGCTTGTGTAGAGAGAGGTCTCCCAAGGCGTCCCGACGCGGC 234  
Qy 188 CGGTAAAACTGGTGGCTGACACCGCGCGGTGGTGAATCGCTATCTCGACATT 247  
Db 293 CCACATGATGCTGTGCGGAGATGACCCGCGGCGACTGATCATCTCGGCTGTAGGT 234  
Qy 248 ACCAAGCGCAGCGCTTGTGTAGCCGACGAGATTGCTGG 283  
Db 233 CCGCAGGTGAGCGGCGCTTGAATTCACAGAGCTGG 198

## RESULT 9

US-09-252-991A-15917  
; Sequence 15917, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15917  
; LENGTH: 480  
; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15917

Query Match 2.8%; Score 36.4; DB 4; Length 480;  
Best Local Similarity 55.6%; Pred. No. 0.12;  
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 254 GCCAGCGTGTAGCCGACGATTTGCTGCGCAAAAAGGCGTCCCGCAGTCTGTCAG 313  
Db 155 GCCAGCGTGTAGCCGCGCGACATCAAGCCGCGCGATTCAGCCGCGCTGATCAG 214  
Qy 314 TCGCATTAATGCTGATGTCGACGACGTCACCGTAAACAGCGGACCTTTCGCCCG 373  
Db 215 TAGCGGTAGGCGCTGCTGCGCAACAGCGCGGAAATCCGCGACAGGCTCCAGCGT 274  
Qy 374 GGCTGG 379  
Db 275 CGCTGG 280

## RESULT 10

US-09-252-991A-15888  
; Sequence 15888, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15888  
; LENGTH: 762  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15888

Query Match 2.8%; Score 36.4; DB 4; Length 762;  
Best Local Similarity 55.6%; Pred. No. 0.16; 56; Indels 0; Gaps 0;  
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 254 GCCAGCGTGTAGCCGACGATTTGCTGCGCAAAAAGGCGTCCCGCAGTCTGTCAG 313  
Db 24 GCCAGCGTGTAGCCGCGCGACATCAAGCCGCGCGATTCAGCCGCGCTGATCAG 83  
Qy 314 TCGCATTAATGCTGATGTCGACGACGTCACCGTAAACAGCGGACCTTTCGCCCG 373  
Db 84 TAGCGGTAGGCGCTGCTGCGCAACAGCGCGGAAATCCGCGACAGGCTCCAGCGT 143  
Qy 374 GGCTGG 379  
Db 144 CGCTGG 149

## RESULT 11

US-09-252-991A-15774/c  
; Sequence 15774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

;; PRIOR FILING DATE: 1998-07-27  
 ;; NUMBER OF SEQ ID NOS: 33142  
 ;; SEQ ID NO 15774  
 ;; LENGTH: 885  
 ;; TYPE: DNA  
 ;; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-15774

Query Match 2.8%; Score 36.4; DB 4; Length 885;  
 Best Local Similarity 55.6%; Pred. No. 0.18; Mismatches 56; Indels 0; Gaps 0;  
 Matches 70; Conservative 0;

QY 254 GCCAGCGCTGTGTAAGCCGAGATTGCTGCGAAAAAGGCGTCCCGCAGTGTGTCAGG 313  
 DB 735 GCCAGCGCTGTGCGCCCGCAGCATCAGCCGCGCGCGATGTCACCGCGCTGATCAGG 676  
 QY 314 TCGCGATTATTGTGTATGTGTCAGAGCCGTACCCGTAAACAGCGGAGCCTTCGCGCGG 373  
 DB 675 TAGCGGTAGGGGCTGTGTGCGCAACAGCGCGGAAATCGCGGACAGGCTCCAGCGTTG 616  
 QY 374 GGCTGG 379  
 DB 615 CGCTGG 610

RESULT 12  
 US-09-252-991A-15804/c  
 ; Sequence 15804, Application US/09252991A  
 ; Patent No. 6551795

;; GENERAL INFORMATION:  
 ;; APPLICANT: Marc J. Rubenfield et al.  
 ;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ;; FILE REFERENCE: 107196.136  
 ;; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ;; CURRENT FILING DATE: 1999-02-18  
 ;; PRIOR APPLICATION NUMBER: US 60/074,788  
 ;; PRIOR FILING DATE: 1998-02-18  
 ;; PRIOR APPLICATION NUMBER: US 60/094,190  
 ;; PRIOR FILING DATE: 1998-07-27  
 ;; NUMBER OF SEQ ID NOS: 33142  
 ;; SEQ ID NO 15804  
 ;; LENGTH: 1068  
 ;; TYPE: DNA  
 ;; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-15804

Query Match 2.8%; Score 36.4; DB 4; Length 1068;  
 Best Local Similarity 55.6%; Pred. No. 0.2; Mismatches 56; Indels 0; Gaps 0;  
 Matches 70; Conservative 0;

QY 254 GCCAGCGCTGTGTAAGCCGAGATTGCTGCGAAAAAGGCGTCCCGCAGTGTGTCAGG 313  
 DB 956 GCCAGCGCTGTGCGCCCGCAGCATCAGCCGCGCGCGATGTCACCGCGCTGATCAGG 897  
 QY 314 TCGCGATTATTGTGTATGTGTCAGAGCCGTACCCGTAAACAGCGGAGCCTTCGCGCGG 373  
 DB 896 TAGCGGTAGGGGCTGTGTGCGCAACAGCGCGGAAATCGCGGACAGGCTCCAGCGTTG 837  
 QY 374 GGCTGG 379  
 DB 836 CGCTGG 831

RESULT 13  
 US-09-710-794-1/c  
 ; Sequence 1, Application US/09710794  
 ; Patent No. 6573069

;; GENERAL INFORMATION:  
 ;; APPLICANT: Holloway, James L.  
 ;; APPLICANT: Gao, Zeren  
 ;; APPLICANT: Whitmore, Theodore F.  
 ;; TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMSEI

;; FILE REFERENCE: 99-76  
 ;; CURRENT APPLICATION NUMBER: US/09/710,794  
 ;; CURRENT FILING DATE: 2000-11-09  
 ;; PRIOR APPLICATION NUMBER: US 60/164,685  
 ;; PRIOR FILING DATE: 1999-11-10  
 ;; NUMBER OF SEQ ID NOS: 31  
 ;; SOFTWARE: FastSeq for Windows Version 3.0  
 ;; SEQ ID NO 1  
 ;; LENGTH: 3076  
 ;; TYPE: DNA  
 ;; ORGANISM: Homo sapiens  
 ;; FEATURE:  
 ;; NAME/KEY: CDS  
 ;; LOCATION: (199)...(1266)  
 US-09-710-794-1

Query Match 2.7%; Score 35.4; DB 4; Length 3076;  
 Best Local Similarity 51.6%; Pred. No. 0.89; Mismatches 76; Indels 0; Gaps 0;  
 Matches 81; Conservative 0;

QY 139 GCCAGCACTGATGACAGATGTCAACCCAGACGATGCGCAACCTTGCGGTAAACTG 198  
 DB 368 GCTTGCTATTGAGAGAGAGGTGTCCCAAGCGTCTCCGCGCGCAAGTGCATG 309  
 QY 199 GATTGCTGACACCGCGGTGTGAGCTAATGCTTATCTCGACATTACCAAGCCAG 258  
 DB 308 GTGTGGCGAAGTGTCCCGCAGCGGCGCTGATCATCTCGCGCTGAGATCGCTCGGGA 249  
 QY 259 GGTCTGTGAGCCGACGAGATTCTGCGCAAAAGGCT 295  
 DB 248 CGGCGCTTGAGTGCACCGAGCTGACACCGATTGCT 212

RESULT 14  
 US-09-902-540-2830  
 ; Sequence 2830, Application US/09902540  
 ; Patent No. 6833447

;; GENERAL INFORMATION:  
 ;; APPLICANT: Goldman, Barry S.  
 ;; APPLICANT: Hinkle, Gregory J.  
 ;; APPLICANT: Slater, Steven C.  
 ;; APPLICANT: Wiegand, Roger C.  
 ;; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ;; FILE REFERENCE: 38-10(15849)B  
 ;; CURRENT APPLICATION NUMBER: US/09/902,540  
 ;; CURRENT FILING DATE: 2001-07-10  
 ;; PRIOR APPLICATION NUMBER: 60/217,883  
 ;; PRIOR FILING DATE: 2000-07-10  
 ;; NUMBER OF SEQ ID NOS: 16825  
 ;; SEQ ID NO 2830  
 ;; LENGTH: 2232  
 ;; TYPE: DNA  
 ;; ORGANISM: Myxococcus xanthus  
 US-09-902-540-2830

Query Match 2.7%; Score 34.6; DB 4; Length 2232;  
 Best Local Similarity 56.6%; Pred. No. 1.3; Mismatches 49; Indels 0; Gaps 0;  
 Matches 64; Conservative 0;

QY 268 GCCAGCACTGCTGTGCGAAAAAGGCGTCCCGCAGTGTGTCAGTGTGCGATTATTGCT 327  
 DB 1294 GCCGCCAAATCTGTGAGAGACCCGGGCGCGCTCGTCCGCTCGCGCGAGCATGCGG 1353  
 QY 328 GATGTGACAGAGGTACCCGTAAACAGGCGAAGCCTTGGCCGCGGCTGGC 380  
 DB 1354 GCGGTGATGATGAGGACCCGAGATGAGAGCGAGAGCTGACGAGAGGAGAC 1406

RESULT 15  
 US-09-902-540-1012/c  
 ; Sequence 1012, Application US/09902540  
 ; Patent No. 6833447  
 ;; GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO. 1012  
LENGTH: 14570  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-1012

Query Match 2.7%; Score 34.6; DB 4; Length 14570;  
Best Local Similarity 56.6%; Pred. No. 4.5;  
Matches 64; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 268 GCCGACGATTCCTGGCGAAGAGGCTCCGCGAGTCTGTCAGTTCGATTTATTCCT 327  
DB 10464 GCCGCCCAATCCTGAGAGACCCGGGGCCGCTCCGTCGCGCGGAGCATGGCG 10405  
QY 328 GATGTCGACGAGCGTACCCGTAACAGCGGAAGCTTCGCCCGCGGGCTGGC 380  
DB 10404 GCGGTCGATGAGCGCACCCGATGAGAGCAGAGACTGGACGACGACGGAC 10352

Search completed: February 10, 2005, 17:43:01  
Job time : 276.446 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 9, 2005, 22:13:58 ; Search time 946.54 Seconds  
(without alignments)  
8097.594 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_12653\_13951

Perfect score: 1299  
Sequence: 1 atgaagaagcattatccccc.....taccgcgttcagttgttaa 1299

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1294.2	99.6	1299	17	US-10-282-122A-7167
2	1293.2	99.6	1901	9	US-09-866-379-7
3	1293.2	99.6	1901	15	US-10-156-660-3
4	1293.2	99.6	1901	17	US-10-601-319-7
5	1291.6	99.4	1901	9	US-09-866-379-5
6	1291.6	99.4	1901	9	US-09-866-379-9
7	1290	99.3	1901	17	US-10-601-319-5
8	1288.4	99.2	1901	9	US-09-866-379-6
9	1288.4	99.2	1901	17	US-10-601-319-6
10	1282.2	98.7	1323	9	US-09-777-566A-1
11	1282.2	98.7	1323	9	US-09-866-379-1

12	1282.2	98.7	1323	14	US-10-034-985-1
13	1282.2	98.7	1323	17	US-10-430-356-1
14	1282.2	98.7	1323	17	US-10-601-319-1
15	1259.2	96.9	1308	15	US-10-156-660-1
16	1259.2	96.9	1308	17	US-10-601-319-9
17	1257.4	96.8	1489	14	US-10-266-041-9
18	1257.4	96.8	1489	17	US-10-284-962-1
19	1252.6	96.4	1486	17	US-10-284-962-4
20	658	51.4	1281	16	US-10-334-672-4
21	668	51.4	1281	16	US-10-334-671-4
22	403.8	31.1	466	17	US-10-282-122A-3181
23	228.6	17.6	1326	15	US-10-021-723A-3
24	224.6	17.3	1325	15	US-10-021-723A-3
25	221.4	17.0	1325	15	US-10-021-723A-1
26	210.8	16.2	1326	15	US-10-317-444-301
27	162.6	12.5	254	17	US-10-317-444-302
28	162.6	12.5	254	17	US-10-317-444-303
29	162.6	12.5	254	17	US-10-317-444-304
30	162.6	12.5	254	17	US-10-021-723A-9
31	157.4	12.1	1431	15	US-10-021-723A-9
32	74	5.7	1230	15	US-10-021-723A-7
33	65.8	5.1	1266	15	US-10-021-723A-5
34	56.2	4.3	11710	16	US-10-240-689-40
35	39.4	3.0	225646	18	US-10-470-565-1
36	39.4	3.0	1068	18	US-10-437-963-53530
37	38	2.9	1395	17	US-10-369-493-41471
38	37.2	2.9	2879	17	US-10-108-260A-2180
39	36	2.8	450	18	US-10-357-930-5437
40	35.6	2.7	427	17	US-10-393-840-372
41	35.4	2.7	344	18	US-10-357-930-4929
42	35.4	2.7	349	18	US-10-357-930-14098
43	35.4	2.7	391	18	US-10-357-930-35227
44	35.4	2.7	391	18	US-10-357-930-44068
45	35.4	2.7	405	18	US-10-357-930-14606

#### ALIGNMENTS

RESULT 1  
US-10-282-122A-7167  
Sequence 7167, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EPIRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625

Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 3181, Ap  
Sequence 41608, A  
Sequence 3, Appli  
Sequence 11, Appli  
Sequence 301, App  
Sequence 302, App  
Sequence 303, App  
Sequence 304, App  
Sequence 9, Appli  
Sequence 7, Appli  
Sequence 5, Appli  
Sequence 40, Appli  
Sequence 53530, A  
Sequence 41471, A  
Sequence 2180, Ap  
Sequence 5437, Ap  
Sequence 372, App  
Sequence 4939, Ap  
Sequence 14098, A  
Sequence 35227, A  
Sequence 44068, A  
Sequence 14606, A



```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7167
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-282-122A-7167

Query Match      99.6%; Score 1294.2; DB 17; Length 1299;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATGAAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTTAAACCCCGCAATCTGCA 60
DB      1 ATGAAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTTAAACCCCGCAATCTGCA 60
QY      61 TTGCGTCAGATGAGCGCGGAGCTGAAAGTGAAGTGTGATTTGTCAGTGTGATGAT 120
DB      61 TTGCGTCAGATGAGCGCGGAGCTGAAAGTGAAGTGTGATTTGTCAGTGTGATGAT 120
QY      121 GTGCGTCTCAACCAAGGCCACGCAACTGATGACAGATGTCAACCCAGACGATGGCCA 180
DB      121 GTGCGTCTCAACCAAGGCCACGCAACTGATGACAGATGTCAACCCAGACGATGGCCA 180
QY      122 GTGCGTCTCAACCAAGGCCACGCAACTGATGACAGATGTCAACCCAGACGATGGCCA 180
DB      122 GTGCGTCTCAACCAAGGCCACGCAACTGATGACAGATGTCAACCCAGACGATGGCCA 180
QY      181 ACCGCGCCGGTAAACTGGGTGGTGGTGAACACCGCGGGGTGGTGAAGTATTCGCTATCTC 240
DB      181 ACCGCGCCGGTAAACTGGGTGGTGGTGAACACCGCGGGGTGGTGAAGTATTCGCTATCTC 240
QY      181 ACCGCGCCGGTAAACTGGGTGGTGGTGAACACCGCGGGGTGGTGAAGTATTCGCTATCTC 240
DB      181 ACCGCGCCGGTAAACTGGGTGGTGGTGAACACCGCGGGGTGGTGAAGTATTCGCTATCTC 240
QY      241 GGACATTAACAACGCGAGCGCTGTGTAGCGCAAGATTCTGCGCAAAAAGGAGCTGCCCG 300
DB      241 GGACATTAACAACGCGAGCGCTGTGTAGCGCAAGATTCTGCGCAAAAAGGAGCTGCCCG 300
QY      301 CAGTCTGATCAGGTGCGATTAATGCTGATGTCGACGACGTAACCCGTAACCAAGCGAA 360
DB      301 CAGTCTGATCAGGTGCGATTAATGCTGATGTCGACGACGTAACCCGTAACCAAGCGAA 360
QY      361 GCGTTGCGCGCGCGGCTGGCACTGCTGCAATTAACCGTACATACCGAGGCAATACG 420
DB      361 GCGTTGCGCGCGCGGCTGGCACTGCTGCAATTAACCGTACATACCGAGGCAATACG 420
QY      421 TCCAGTCCCGATCCGTTATTTATCTCTAATAAACTGGGTTTGCACCTGATTAACGCG 480
DB      421 TCCAGTCCCGATCCGTTATTTATCTCTAATAAACTGGGTTTGCACCTGATTAACGCG 480
QY      481 AACGTGACTGACGCGATCTCTACAGAGGCAAGAGGCTCAATTGCTGACTTTAACCGGCA 540
DB      481 AACGTGACTGACGCGATCTCTACAGAGGCAAGAGGCTCAATTGCTGACTTTAACCGGCA 540
QY      541 CCGCAACCGCGCTTTGGCGCAACTGGAACCGGCTTTAATTTTCCGAATCAAACTTTGCG 600
DB      541 CCGCAACCGCGCTTTGGCGCAACTGGAACCGGCTTTAATTTTCCGAATCAAACTTTGCG 600
QY      601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGCAATTAACATGGAAC 660
DB      601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGCAATTAACATGGAAC 660
QY      661 AAGGTAGCGCGCAATGCTCATTTAACCGGTGCGGTAAAGCTCCCATCAATGCTGAGCG 720
DB      661 AAGGTAGCGCGCAATGCTCATTTAACCGGTGCGGTAAAGCTCCCATCAATGCTGAGCG 720
QY      721 GAGATTTTCTCTGCAACAGACAGGGAATGCCGAGCGCGGGGTGGGGAAGAGATCAC 780
DB      721 GAGATTTTCTCTGCAACAGACAGGGAATGCCGAGCGCGGGGTGGGGAAGAGATCAC 780
QY      781 GATTACACCAAGTGAACACTTGTGAAGTTTGCATTAACGCGCAATTTATTTGCTACAA 840
DB      781 GATTACACCAAGTGAACACTTGTGAAGTTTGCATTAACGCGCAATTTATTTGCTACAA 840
```

```

DB      781 GATTACACCAAGTGAACACTTGTGAAGTTTGCATTAACGCGCAATTTATTTGCTACAA 840
QY      841 CGCACCGCAGAGGTTTGGCCGAGCGCGCCACCCCGTTATTAAGTTGATTAACAGCG 900
DB      841 CGCACCGCAGAGGTTTGGCCGAGCGCGCCACCCCGTTATTAAGTTGATTAACAGCG 900
QY      901 TTGACGCCCATTCACCGCAAAAACAGCGGTATGATGATTAACCACTTCACTGCTG 960
DB      901 TTGACGCCCATTCACCGCAAAAACAGCGGTATGATGATTAACCACTTCACTGCTG 960
QY      961 TTTATGCGCGGACACAGATTAATCTGCAATCTCGCGGCGCACTGAGCTCACTG 1020
DB      961 TTTATGCGCGGACACAGATTAATCTGCAATCTCGCGGCGCACTGAGCTCACTG 1020
QY      1021 ACGCTTCCCGGTCAAGCCGATTAACAGCGCGCGAGTGTGAACGTGTTTGAACGCTG 1080
DB      1021 ACGCTTCCCGGTCAAGCCGATTAACAGCGCGCGAGTGTGAACGTGTTTGAACGCTG 1080
QY      1081 CGTGCGCTAAGCGATTAACAGCGAGTGAATGAGTTTGTGCTGTGCTTCCAGACTTTAC 1140
DB      1081 CGTGCGCTAAGCGATTAACAGCGAGTGAATGAGTTTGTGCTGTGCTTCCAGACTTTAC 1140
QY      1141 CAGATCGTGAATAAACCGCGCTGTCAATTAATAACCGCGCGGAGAGTGAACCTGACC 1200
DB      1141 CAGATCGTGAATAAACCGCGCTGTCAATTAATAACCGCGCGGAGAGTGAACCTGACC 1200
QY      1201 CTGGCAGATGATGAAGACGAAATGCGCAGGCGAGTGTGTTGTTGAGGTTTAAACG 1260
DB      1201 CTGGCAGATGATGAAGACGAAATGCGCAGGCGAGTGTGTTGTTGAGGTTTAAACG 1260
QY      1261 ATCGTGAATGAAGACGATACCGCGCTTGACATTTGTTAA 1299
DB      1261 ATCGTGAATGAAGACGATACCGCGCTTGACATTTGTTAA 1299

RESULT 2
US-09-866-379-7
; Sequence 7, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KREITZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BAYRON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1901)
; OTHER INFORMATION: n is any nucleotide
US-09-866-379-7
```

Query Match 99.6%; Score 1293.2; DB 9; Length 1901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTGATTCGGTTAACCCCGAATCTGCA 60  
188 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTGATTCGGTTAACCCCGAATCTGCA 247  
61 TTGCGCTCAGATGAGCCGAGCTGAAAGTGTGATGATGTCAGTCTCATGCT 120  
248 TTGCGCTCAGATGAGCCGAGCTGAAAGTGTGATGATGTCAGTCTCATGCT 307  
121 GTGCGCTCTCAACCAAGGCCAAGCACTGATGAGATGTCACCCCAAGCCGATGCCA 180  
308 GTGCGCTCTCAACCAAGGCCAAGCACTGATGAGATGTCACCCCAAGCCGATGCCA 367  
181 ACCTGGCCGGTAAATCTGGGTTGGCTGACACCGCGGGTGGAGCTTAATCGCTATCTC 240  
368 ACCTGGCCGGTAAATCTGGGTTGGCTGACACCGCGGGTGGAGCTTAATCGCTATCTC 427  
241 GGACATTAACAACGCCAGGCTGTGATGCGAGCAATGCTGGCAAAAAGGCTGCCG 300  
428 GGACATTAACAACGCCAGGCTGTGATGCGAGCAATGCTGGCAAAAAGGCTGCCG 487  
301 CAGCTGTGATCAGGTGCGCATTTATGCTGATGTCAGACGCGTAACCCGTAAACAGCGCA 360  
488 CAGCTGTGATCAGGTGCGCATTTATGCTGATGTCAGACGCGTAACCCGTAAACAGCGCA 547  
361 GCGTTTGGCGCGCGGCTGCGCACTGACTGTGCAATAACCTAATACCCGAGCAATAC 420  
548 GCGTTTGGCGCGCGGCTGCGCACTGACTGTGCAATAACCTAATACCCGAGCAATAC 607  
421 TCCAGTCCCGATCGTTATTTAATCTCTTAAACTGCGCTTGGCAACTGATTAACGCG 480  
608 TCCAGTCCCGATCGTTATTTAATCTCTTAAACTGCGCTTGGCAACTGATTAACGCG 667  
481 AACGTGACTGACGCGATCTCTCAGCAGGCGAGAGGGTCAATGCTGACTTAACCGGCGAT 540  
668 AACGTGACTGACGCGATCTCTCAGCAGGCGAGAGGGTCAATGCTGACTTAACCGGCGAT 727  
541 CCGCAAAACGGCGTTTGGCGCACTGGAACGGGTGCTTAATTTTCCGCAATCAACTGTGC 600  
728 CCGCAAAACGGCGTTTGGCGCACTGGAACGGGTGCTTAATTTTCCGCAATCAACTGTGC 787  
601 CTTAAACGTGAGAAACGAGCAAGAAAGCTGTTCAATTACGAGGCAATTAACATCGAATC 660  
788 CTTAAACGTGAGAAACGAGCAAGAAAGCTGTTCAATTACGAGGCAATTAACATCGAATC 847  
661 AAGGTGAGCGCGCAATGCTCTCAATTAACCGGTGCGTAAGCTCGCATCAATGCTGAGC 720  
848 AAGGTGAGCGCGCAATGCTCTCAATTAACCGGTGCGTAAGCTCGCATCAATGCTGAGC 907  
721 GAGATATTTTCTCTGCAACAGCAAGGAAATGCGGAGCGCGGGTGGGGAAGATAC 780  
908 GAGATATTTTCTCTGCAACAGCAAGGAAATGCGGAGCGCGGGTGGGGAAGATAC 967  
781 GATTCAACAGGAGGAAACCTTGTCTAAGTTTGATTAAGCGCAATTTTAATTTGCTACAA 840  
968 GATTCAACAGGAGGAAACCTTGTCTAAGTTTGATTAAGCGCAATTTTAATTTGCTACAA 1027  
841 CGCAGCGCAGAGGTGGCCGAGCGCGCAACCCCGTTAATTAAGTTTGAATCAAGACGCG 900  
1028 CGCAGCGCAGAGGTGGCCGAGCGCGCAACCCCGTTAATTAAGTTTGAATCAAGACGCG 1087  
901 TTGACGCGCCCATCCACCGCAAAAACAGCGCTTAATGAGTGAATTAACCACTTCAGTGTG 960  
1088 TTGACGCGCCCATCCACCGCAAAAACAGCGCTTAATGAGTGAATTAACCACTTCAGTGTG 1147  
961 TTATATGCGCGAGACGATTAATCTGAGCAAAATCTCGCGCGCGCATGAGACTCAACTG 1020  
1148 TTATATGCGCGAGACGATTAATCTGAGCAAAATCTCGCGCGCGCATGAGACTCAACTG 1207

1021 ACGTTCCCGGTCAGCGGATTAACAGCGCGCAGGTGATGAACTGTGTTGAACGCTGG 1080  
1208 ACGTTCCCGGTCAGCGGATTAACAGCGCGCAGGTGATGAACTGTGTTGAACGCTGG 1267  
1081 CGTGGCTTAAGGATTAACAGCGGATGATGATGATGATGATGATGATGATGATGATG 1140  
1268 CGTGGCTTAAGGATTAACAGCGGATGATGATGATGATGATGATGATGATGATGATG 1327  
1141 CAGATGCGTGAATTAACCGCGCTGCTCAATTAATAGCCCGCGGAGAGGTGAACGAGCC 1200  
1328 CAGATGCGTGAATTAACCGCGCTGCTCAATTAATAGCCCGCGGAGAGGTGAACGAGCC 1387  
1201 CTGGCAGATGATGAAGAGGAAATGCGCAGGCAATGATGATGATGATGATGATGATG 1260  
1388 CTGGCAGATGATGAAGAGGAAATGCGCAGGCAATGATGATGATGATGATGATGATG 1447  
1261 ATCGTGAATGAAGACGATTAACCGCTTGCAGTTGTAA 1299  
1448 ATCGTGAATGAAGACGATTAACCGCTTGCAGTTGTAA 1486

RESULT 3  
US-10-156-660-3  
; Sequence 3, Application US/1015660  
; Publication No. US20030103958A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; APPLICANT: Kretz, Keith  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Barton, Nelson R.  
; APPLICANT: Garrett, James B.  
; APPLICANT: O'Donoghue, Eileen  
; TITLE OF INVENTION: PHYLASES, NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 09010-029007  
; CURRENT APPLICATION NUMBER: US/10/156,660  
; CURRENT FILING DATE: 2002-10-01  
; PRIOR APPLICATION NUMBER: US 09/866,379  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1901  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (188) ... (1483)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 403  
; OTHER INFORMATION: n = A,T,C or G  
US-10-156-660-3

Query Match 99.6%; Score 1293.2; DB 15; Length 1901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTGATTCGGTTAACCCCGAATCTGCA 60  
188 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTGATTCGGTTAACCCCGAATCTGCA 247

QY 61 TTCCGCTCAGAGTGAGCCCGGAGCTGAAGCTGGAAGTGTGGTATTTGTCACTGCTATGCT 120  
DB 248 TTCCGCTCAGAGTGAGCCCGGAGCTGAAGCTGGAAGTGTGGTATTTGTCACTGCTATGCT 307  
QY 121 GTGGCGTGTCCAAACCAAGGCGGACGCAACGTATGAGATGTCACCCGACGATGGCGCA 180  
DB 308 GTGGCGTGTCCAAACCAAGGCGGACGCAACGTATGAGATGTCACCCGACGATGGCGCA 367  
QY 181 ACCGCGCGGTAAACCTGAGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240  
DB 368 ACCGCGCGGTAAACCTGAGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 427  
QY 241 GGAATTTACCAACGCGGACGCTGTGATGCGGATGCTGCGGCAAAAAGGCGTGGCCG 300  
DB 428 GGAATTTACCAACGCGGACGCTGTGATGCGGATGCTGCGGCAAAAAGGCGTGGCCG 487  
QY 301 CAGTCTGGTCAAGTGTGCGGATTTATGCTGATGTCAGACGCGTAAACCGGACGCA 360  
DB 488 CAGTCTGGTCAAGTGTGCGGATTTATGCTGATGTCAGACGCGTAAACCGGACGCA 547  
QY 361 GCGTTTCCGCGCGGCGCTGGACCTGATGTCATTAACCGTAACTAATCCGACGATACG 420  
DB 548 GCGTTTCCGCGCGGCGCTGGACCTGATGTCATTAACCGTAACTAATCCGACGATACG 607  
QY 421 TCCAGTCCCGATCCGTTATTTAATCTCTAAACCTGGCGTTGGCACTGGATTAACGCG 480  
DB 608 TCCAGTCCCGATCCGTTATTTAATCTCTAAACCTGGCGTTGGCACTGGATTAACGCG 667  
QY 481 AACGTGACTGACGCGATCTCTAGACGAGGAGGAGGATGCTGACTTTTACCGGCGAT 540  
DB 668 AACGTGACTGACGCGATCTCTAGACGAGGAGGAGGATGCTGACTTTTACCGGCGAT 727  
QY 541 CGGCAAAACGCGGTTTTCGCACTGGAACGCGGTCTTTAATTTTCCGCAATCAACTTGTGC 600  
DB 728 CGGCAAAACGCGGTTTTCGCACTGGAACGCGGTCTTTAATTTTCCGCAATCAACTTGTGC 787  
QY 601 CTTAAACGTGAAGAACAGGACGAAAGCTGTATTAAGCAGGCACTTACATCGGAATC 660  
DB 788 CTTAAACGTGAAGAACAGGACGAAAGCTGTATTAAGCAGGCACTTACATCGGAATC 847  
QY 661 AAGTGAGCGCGGACCAATGTCTCAATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 720  
DB 848 AAGTGAGCGCGGACCAATGTCTCAATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 907  
QY 721 GAGATATTTTCTCTGCAACAGACAGGAAATGCGGAGCCGCGGTGTGGGAAAGATCAAC 780  
DB 908 GAGATATTTTCTCTGCAACAGACAGGAAATGCGGAGCCGCGGTGTGGGAAAGATCAAC 967  
QY 781 GATTCAACACAGTGGGAACCTTGTCTAATGTTGCAATTAAGCGCAATTTATTTGCTACAA 840  
DB 968 GATTCAACACAGTGGGAACCTTGTCTAATGTTGCAATTAAGCGCAATTTATTTGCTACAA 1027  
QY 841 CGCAGCGCAGAGGTTGCGCGGACGCGGACCGCGCTTTAATGATTTGATCAAGACAGCG 900  
DB 1028 CGCAGCGCAGAGGTTGCGCGGACGCGGACCGCGCTTTAATGATTTGATCAAGACAGCG 1087  
QY 901 TTGAGCGCCCATTCACCCGCAAAAACGCGGTATGATGATGATTAACCTTCAAGTGTG 960  
DB 1088 TTGAGCGCCCATTCACCCGCAAAAACGCGGTATGATGATGATTAACCTTCAAGTGTG 1147  
QY 961 TTTATGCGCGGACAGATCTAATCTGGAATCTGCGGCGGACATGAGAGCTCAACTG 1020  
DB 1148 TTTATGCGCGGACAGATCTAATCTGGAATCTGCGGCGGACATGAGAGCTCAACTG 1207  
QY 1021 ACGCTTCCCGGTTCAGCGCGATTAACGCGCGCGGATGATGAATGATGTTTGAACGCTGG 1080  
DB 1208 ACGCTTCCCGGTTCAGCGCGATTAACGCGCGCGGATGATGAATGATGTTTGAACGCTGG 1267  
QY 1081 CGTGGCTTAAGGATTAACAGCGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
DB 1268 CGTGGCTTAAGGATTAACAGCGATGATGATGATGATGATGATGATGATGATGATGATG 1327  
QY 1141 CAGATGCGTGAATAAACGCGGTGTGATTAATAGCGCGCGGATGAGATGAACCTGACG 1200

DB 1328 CAGATGCGTGAATAAACGCGGTGTGATTAATAGCGCGCGGATGAGATGAACCTGACG 1387  
QY 1201 CTGGCAGGATGTAAGACGCAAAATGCGGAGGCAATGTTGTTGCGAGGTTTAAACGCA 1280  
DB 1388 CTGGCAGGATGTAAGACGCAAAATGCGGAGGCAATGTTGTTGCGAGGTTTAAACGCA 1447  
QY 1281 ATCGGAATGAAGACGATACCGGCTTGGCATTTGTA 1299  
DB 1448 ATCGGAATGAAGACGATACCGGCTTGGCATTTGTA 1486  
RESULT 4  
US-10-601-319-7  
; Sequence 7, Application US/10601319  
; Publication No. US20040091968A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; APPLICANT: Kretz, Keith A.  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Barton, Nelson Robert  
; APPLICANT: Garrett, James B.  
; APPLICANT: O' Donoghue, Eileen  
; APPLICANT: Machuz, Eric J.  
; TITLE OF INVENTION: RECOMBINANT PHYTASES AND METHODS OF MAKING  
; TITLE OF INVENTION: AND USING THEM  
; FILE REFERENCE: 09010-029011  
; CURRENT APPLICATION NUMBER: US/10/601,319  
; PRIOR FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: US 09/866,379  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 1901  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (188)...(1483)  
; REFEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 403  
; OTHER INFORMATION: n = A,T,C or G  
; US-10-601-319-7  
Query Match 99.6%; Score 1293.2; DB 17; Length 1901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1995; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ATGAAGCAGATCTTAATCCATTTTATCTCTTGTGATTCGGTTAAACCCGCAATCTGCA 60  
DB 188 ATGAAGCAGATCTTAATCCATTTTATCTCTTGTGATTCGGTTAAACCCGCAATCTGCA 247  
QY 61 TTCCGCTCAGAGTGAGCCCGGAGCTGAAGCTGGAAGTGTGGTATTTGTCACTGCTATGCT 120  
DB 248 TTCCGCTCAGAGTGAGCCCGGAGCTGAAGCTGGAAGTGTGGTATTTGTCACTGCTATGCT 307  
QY 121 GTGGCGTGTCCAAACCAAGGCGGACGCAACGTATGAGATGTCACCCGACGATGGCGCA 180  
DB 308 GTGGCGTGTCCAAACCAAGGCGGACGCAACGTATGAGATGTCACCCGACGATGGCGCA 367  
QY 181 ACCGCGCGGTAAACCTGAGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240  
DB 368 ACCGCGCGGTAAACCTGAGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC

Db	368	ACCTGGCCGGTAAAACTGGGTTGGCTGACACCGCGAGGTGTGAGCTTAATCGCTTATCTTC	427
Qy	241	GGACATTACAAACGCGCAGCGTCTGTGTAGCCGACGAGATTGCTGGCGAAAAAGGGCTGCCG	300
Db	428	GGACATTACAAACGCGCGCGCTGTGTAGCCGACGAGATTGCTGGCGAAAAAGGGCTGCCG	487
Qy	301	CAGTCTGTGTCAGGTGCGGATTTATTGCTGATGTGACGAGCTTACCCGTAAACAGCGGAA	360
Db	488	CAGTCTGTGTCAGGTGCGGATTTATTGCTGATGTGACGAGCTTACCCGTAAACAGCGGAA	547
Qy	361	GCCTTCCCGCCGGGGCTGGGACCGTGTGTGCAATTAACGTAATATCCAGGACGATACG	420
Db	548	GCCTTCCCGCCGGGGCTGGGACCGTGTGTGCAATTAACGTAATATCCAGGACGATACG	607
Qy	421	TCCAGTCCCGATCCGTTATTTAACTCTTAAAACTGGCGTTTGGCAACTGTGATTAACGCG	480
Db	608	TCCAGTCCCGATCCGTTATTTAACTCTTAAAACTGGCGTTTGGCAACTGTGATTAACGCG	667
Qy	481	AACGTGACTGACGGGATCTTCAACGAGGACGAGGAGGTCATTTGCTGACTTTACCGGCGAT	540
Db	668	AACGTGACTGACGGGATCTTCAACGAGGACGAGGAGGTCATTTGCTGACTTTACCGGCGAT	727
Qy	541	CGGCAAAACGGCGTTTCCGGAATCTGGAACGGGTGTCTTAATTTCCGCAATCAACTGTGCG	600
Db	728	CGGCAAAACGGCGTTTCCGGAATCTGGAACGGGTGTCTTAATTTCCGCAATCAACTGTGCG	787
Qy	601	CTTAAACGTGAGAAACAGAGCAAGAAAGCTGTTCATTAAACGACGACATTACCATCGAATCC	660
Db	788	CTTAAACGTGAGAAACAGAGCAAGAAAGCTGTTCATTAAACGACGACATTACCATCGAATCC	847
Qy	661	AAGGTGAGCGCGCAACAATGTCTCATTTAACCGGTGCGGTAAAGCTCGATCAATGTCTGACG	720
Db	848	AAGGTGAGCGCGCAACAATGTCTCATTTAACCGGTGCGGTAAAGCTCGATCAATGTCTGACG	907
Qy	721	GAGATATTTCTCTCTGCAACAGCAACAGGAATGCCCGGACCGGGGTGGGGAAGGATCAC	780
Db	908	GAGATATTTCTCTCTGCAACAGCAACAGGAATGCCCGGACCGGGGTGGGGAAGGATCAC	967
Qy	781	GATTACACACAGTGGGACACCTTGCTAAGTTTGATTAACGCGCAATTTATTTTGCTTCAAA	840
Db	968	GATTACACACAGTGGGACACCTTGCTAAGTTTGATTAACGCGCAATTTATTTTGCTTCAAA	1027
Qy	841	CGCACGCGCAAGGTTGCCCGAGCCGCGCCACCCCGTTATTAGATTGGATCAGACAGCG	900
Db	1028	CGCACGCGCAAGGTTGCCCGAGCCGCGCCACCCCGTTATTAGATTGGATCAGACAGCG	1087
Qy	901	TTGACGCGCCATCCACCGCGAAAAACAGCGGTATGTGTGACATTAACCATCTTCAGTCTG	960
Db	1088	TTGACGCGCCATCCACCGCGAAAAACAGCGGTATGTGTGACATTAACCATCTTCAGTCTG	1147
Qy	961	TTTATCGCGCGGACAGATTAATTAATGTGGCAATCTCGCGCGCGACAGTGGAGCTCAACTGG	1020
Db	1148	TTTATCGCGCGGACAGATTAATTAATGTGGCAATCTCGCGCGCGACAGTGGAGCTCAACTGG	1207
Qy	1021	ACGCTTCCCGGTCAACCGGATTAACGCGCGCCAGGCTGTGTGAATGTGTTTGAACGCTGG	1080
Db	1208	ACGCTTCCCGGTCAACCGGATTAACGCGCGCCAGGCTGTGTGAATGTGTTTGAACGCTGG	1267
Qy	1081	CGTCCGCTTAAGCGATTAACGCGCGCTGCATTAAATACGCGCCCGGAGAGGTGAACCTAGC	1140
Db	1268	CGTCCGCTTAAGCGATTAACGCGCGCTGCATTAAATACGCGCCCGGAGAGGTGAACCTAGC	1327
Qy	1141	CAGATGCGGATTAATAACGCGCGCTGCATTAAATACGCGCCCGGAGAGGTGAACCTAGC	1200
Db	1328	CAGATGCGGATTAATAACGCGCGCTGCATTAAATACGCGCCCGGAGAGGTGAACCTAGC	1387
Qy	1201	CTGGCAGGATGTGAAGAGGGAATCGCAGGGCATGTGTCGTTGGCAGGTTTACGCAA	1260
Db	1388	CTGGCAGGATGTGAAGAGGGAATCGCAGGGCATGTGTCGTTGGCAGGTTTACGCAA	1447
Qy	1261	ATCGTGAATGAAGACGCAATACCGCGTTGCAATTTGTAA	1299

Db	1448 ATCGTGAATGAGACCGCATACCGGCGCTGCAGTTGTGTA	1486
Query Match	99.4%; Score 1291.6; DB 9; Length 1901;	
Best Local Similarity	99.6%; Fred. No. 0;	
Matches 1294; Conservative	0; Mismatches 5; Indels 0; Gaps 0	
Query	1 ATGAAGCATCTTATCCCAATTTTATCTCTTCGATTCGTTAACCCCGCATCTCA	60
Db	168 ATGAAGCATCTTATCCCAATTTTATCTCTTCGATTCGTTAACCCCGCATCTCA	247
Query	61 TTCGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCATGTCATGCT	120
Db	248 TTCGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCATGTCATGCT	307
Query	121 GTGCGTGTCTCCAAACCAAGGCCCAAGTATGACAGATGACCCCAAGACGATGGACA	180
Db	308 GTGCGTGTCTCCAAACCAAGGCCCAAGTATGACAGATGACCCCAAGACGATGGACA	367
Query	181 AACTGACGCGTAAACTGGGTTGCTGACACCGCCGCGTGTGATGACTATCGCTTATCTC	240
Db	368 AACTGACGCGTAAACTGGGTTGCTGACACCGCGGCGTGTGATGACTATCGCTTATCTC	427
Query	241 GGAACATTAACAACGCGCAGCGTGTGTAAGCCGACGATTTGTCGGAAAAAGGCTGCCCG	300
Db	428 GGAACATTAACAACGCGCAGCGTGTGTAAGCCGACGATTTGTCGGAAAAAGGCTGCCCG	487
Query	301 CAGTGTGTCAGAGTGCAGTATTAATGTCATGTCGACGAGCGTAAACCGTAAACAGCGCA	360
Db	488 CAGTGTGTCAGAGTGCAGTATTAATGTCATGTCGACGAGCGTAAACCGTAAACAGCGCA	547
Query	361 GCCTTGCAGCCGCGGCTGCGACCTGACTGTGTGCATTAACGTAATACCCAGGCAATACG	420
Db	548 GCCTTGCAGCCGCGGCTGCGACCTGACTGTGTGCATTAACGTAATACCCAGGCAATACG	607

QY	421	TCAGTCCCGATCCGGTATTTAATCCCTTAAATACGCGCTTGCCAACTGGATTACGGG	480
Dp	608	TCAGTCCCGATCCGGTATTTAATCCCTTAAATACGCGCTTGCCAACTGGATTACGGG	667
QY	481	AACGTGACTGACCGGATCTCAGCAGGGCAGAGGGTCAATTGCTGACTTTACCGGCAT	540
Dp	668	AACGTGACTGACCGGATCTCAGCAGGGCAGAGGGTCAATTGCTGACTTTACCGGCAT	727
QY	541	CGGCAAAACGGCGTTTCGCGCAACTGGAAACGGGTGTTAATTTTCGCAATCAAACTTGTC	600
Dp	728	CGGCAAAACGGCGTTTCGCGCAACTGGAAACGGGTGTTAATTTTCGCAATCAAACTTGTC	787
QY	601	CTTAAACGTGAGAAACAGACGAAAGCTGTTCATTAAACGAGGCATTACCATCGAACTC	660
Dp	788	CTTAAACGTGAGAAACAGACGAAAGCTGTTCATTAAACGAGGCATTACCATCGAACTC	847
QY	661	AAGGTGACCGCGCAATGTCTCATTTAACCGGTGCGGTAAAGCTCGCATCATAGTCTGACG	720
Dp	848	AAGGTGACCGCGCAATGTCTCATTTAACCGGTGCGGTAAAGCTCGCATCATAGTCTGACG	907
QY	721	GAGATATTTCTCCGCAAAACAGACACAGGAAATGCCGAGCCGGGGTGGGGAAGATCAC	780
Dp	908	GAGATATTTCTCCGCAAAACAGACACAGGAAATGCCGAGCCGGGGTGGGGAAGATCAC	967
QY	781	GATTCAACACAGTGGAAACACCTTGCTAAGTTTGCAATAACCGCAATTTATTTTGCTACAA	840
Dp	968	GATTCAACACAGTGGAAACACCTTGCTAAGTTTGCAATAACCGCAATTTATTTTGCTACAA	1027
QY	841	CGCACGCCAGAGGTGGCCCGGACGCCGCCACCCCGTATTAAAGTTTGAACAAGACACG	900
Dp	1028	CGCACGCCAGAGGTGGCCCGGACGCCGCCACCCCGTATTAAAGTTTGAACAAGACACG	1087
QY	901	TTGAACGCCCATCCACCGCAAAAACAGGCGTATGTTGACATTAACCACTTGACGTGCTG	960
Dp	1088	TTGAACGCCCATCCACCGCAAAAACAGGCGTATGTTGACATTAACCACTTGACGTGCTG	1147
QY	961	TTTATCGCCGACACGATACTTAATCTGGCAATCTCGCGCGCGACCTGGAGCTCACTGG	1020
Dp	1148	TTTATCGCCGACACGATACTTAATCTGGCAATCTCGCGCGCGACCTGGAGCTCACTGG	1207
QY	1021	ACGCTTCCCGGTACGCGGATACACGCGCGCACGATGATGAATCTGATTTGAACGCTGG	1080
Dp	1208	ACGCTTCCCGGTACGCGGATACACGCGCGCACGATGATGAATCTGATTTGAACGCTGG	1267
QY	1081	CGTGGCTAAGCGATTAACGCGCAGTGAATTCAGGTTTGGCTGTCTTCAGACTTTACAG	1140
Dp	1268	CGTGGCTAAGCGATTAACGCGCAGTGAATTCAGGTTTGGCTGTCTTCAGACTTTACAG	1327
QY	1141	CAGATGCGGTAAACGCGCGCTGCATTAAATACGCGCCCGGAGAGGTGAACCTAC	1200
Dp	1328	CAGATGCGGTAAACGCGCGCTGCATTAAATACGCGCCCGGAGAGGTGAACCTAC	1387
QY	1201	CTGGCAGGATGTGAAGACGAAATGCGCAGGGCAATGTGTTCGTTGGCAGGTTTACGCAA	1260
Dp	1388	CTGGCAGGATGTGAAGACGAAATGCGCAGGGCAATGTGTTCGTTGGCAGGTTTACGCAA	1447
QY	1261	ATCTGAAATGAAGCACGATACCGGCTTGCACTTTGTA	1299
Dp	1448	ATCTGAAATGAAGCACGATACCGGCTTGCACTTTGTA	1486

RESULT 6  
US-09-866-379-9  
Sequence 9, Application US/09866379  
Patent No. US20020136754A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
INVENTOR: SHORT, Jay  
APPLICANT: KRETTZ, Keith  
APPLICANT: GRAY, Kevin  
APPLICANT: BARTON, Nelson  
APPLICANT: GARRETT, James  
APPLICANT: O'DONOGHUE, Eileen

```

/ TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
/ FILE REFERENCE: DIVER1370-7
/ CURRENT APPLICATION NUMBER: US/09/866,379
/ CURRENT FILING DATE: 2001-05-24
/ PRIOR APPLICATION NUMBER: US 09/580,515
/ PRIOR FILING DATE: 2000-05-25
/ PRIOR APPLICATION NUMBER: US 09/318,528
/ PRIOR FILING DATE: 1999-05-25
/ PRIOR APPLICATION NUMBER: US 09/291,931
/ PRIOR FILING DATE: 1999-04-13
/ PRIOR APPLICATION NUMBER: US 09/259,214
/ PRIOR FILING DATE: 1999-03-01
/ PRIOR APPLICATION NUMBER: US 08/910,758
/ PRIOR FILING DATE: 1997-08-13
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 9
/ LENGTH: 1901
/ TYPE: DNA
/ ORGANISM: Escherichia coli
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1)..(1901)
/ OTHER INFORMATION: n is any nucleotide
US-09-866-379-9

```

Query Match	99.4%;	Score 1291.6;	DB 9;	Length 1901;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1294;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;
QY	1	ATGAAGCATCTTAATCCCATTTTAACTCTTCGTGATTCGCTTAACCCCGCAATCTGCA	60	
Db	188	ATGAAGGAGATCTTAATCCCATTTTAACTCTTCGTATTCGGTTAACCCCGCAATCTGCA	247	
QY	61	TTGCGTCAGAGTGAGCCGGAGCTGAAGTGAAGAGTGATGTTGCACTGCTCATGCT	120	
Db	248	TTGCGTCAGAGTGAGCCGGAGCTGAAGTGAAGAGTGATGTTGCACTGCTCATGCT	307	
QY	121	GTGCGTGTCTCAACCAAGGCCACGCACTGATGCAAGATGTCACCCGAGAGCATGGCCA	180	
Db	308	GTGCGTGTCTCAACCAAGGCCACGCACTGATGCAAGATGTCACCCGAGAGCATGGCCA	367	
QY	181	AACGGCCCGGTAAAACTGGATTGGGTGACACCGCGCGGTGGTGAAGCTATCGCCTATCTC	240	
Db	368	AACGGCCCGGTAAAACTGGATTGGGTGACACCGCGNGGTGGTGAAGCTATCGCCTATCTC	427	
QY	241	GGACATTACCAACGCCACAGCTGTGGTAGCCGACGCAATTGCTGGCGAAAAAGGGCTGGCCG	300	
Db	428	GGACATTACCAACGCCACAGCTGTGGTAGCCGACGCAATTGCTGGCGAAAAAGGGCTGGCCG	487	
QY	301	CAGTCTGGTCAAGTGCAGCTAATTGTCGTATGTCGACGAGCGTACCCGTTAAAAACAGCGCA	360	
Db	488	CAGTCTGGTCAAGTGCAGCTAATTGTCGTATGTCGACGAGCGTACCCGTTAAAAACAGCGCA	547	
QY	361	GCCTTGGCGCGGGCTGGCACTGATGTCGCAATTAACGTAACATACCAGGACAGATACG	420	
Db	548	GCCTTGGCGCGGGCTGGCACTGATGTCGCAATTAACGTAACATACCAGGACAGATACG	607	
QY	421	TCCAGTCCCGGATCCGTTATTTAACTCTCTAATAAACTGGCGTTTGGCAACTGGATTAACCG	480	
Db	608	TCCAGTCCCGGATCCGTTATTTAACTCTCTAATAAACTGGCGTTTGGCAACTGGATTAACCG	667	
QY	481	AAACGTGACTGACGCGATCTTACGACGAGGACGAGAGGTCATATTGCTGACTTTAACCGGCGAT	540	
Db	668	AAACGTGACTGACGCGATCTTACGACGAGGACGAGAGGTCATATTGCTGACTTTAACCGGCGAT	727	
QY	541	CGGCAAAAGGGGCTTTCGGGAACTGGAAAGGGGAGCTAATTTTTCGGCAATCAAACCTGTGTC	600	
Db	728	CGGCAAAAGGGGCTTTCGGGAACTGGAAAGGGGAGCTAATTTTTCGGCAATCAAACCTGTGTC	787	
QY	601	CTTAAACGTGAGAAACAGAGCAAAAGCTGTTCATTAACGACAGGCACTTAACATCGGAACTC	660	
Db	788	CTTAAACGTGAGAAACAGAGCAAAAGCTGTTCATTAACGACAGGCACTTAACATCGGAACTC	847	

QY 661 AAGGTAGGCGCGACATGTCATTACCGGTCGTAAGCTCGCATCAATGCTGACG 720  
DB 848 AAGGTAGGCGCGACATGTCATTACCGGTCGTAAGCTCGCATCAATGCTGACG 907  
QY 721 GAGATATTTCTCTGCAACAAGCAAGGGAATGCGGAGCGGGGTGGGGGAAGATCACC 780  
DB 908 GAGATATTTCTCTGCAACAAGCAAGGGAATGCGGAGCGGGGTGGGGGAAGATCACC 967  
QY 781 GATTACACCAAGTGAACAACCTTGTATAGTTGCATTAACGCGAATTTTATTGCTACAA 840  
DB 968 GATTACACCAAGTGAACAACCTTGTATAGTTGCATTAACGCGAATTTTATTGCTACAA 1027  
QY 841 CGCAGCGCGAGAGTTGCGCGACCGCGCCACCCCGTTATTAGTTGATCAAGACGCG 900  
DB 1028 CGCAGCGCGAGAGTTGCGCGACCGCGCCACCCCGTTATTAGTTGATCAAGACGCG 1087  
QY 901 TTGACCGCCCATCCACCGCAAAAACAGGCGTATGTTGACATTAACCACTTCAAGTCTG 960  
DB 1088 TTGACCGCCCATCCACCGCAAAAACAGGCGTATGTTGACATTAACCACTTCAAGTCTG 1147  
QY 961 TTATGCGCGGACGATTAATCTGGCAAAATCTCGCGCGCGCATGAGACTCAACTGCG 1020  
DB 1148 TTATGCGCGGACGATTAATCTGGCAAAATCTCGCGCGCGCATGAGACTCAACTGCG 1207  
QY 1021 AGCGTTCCCGGTACAGCGGATTAACGCGCGCGAGGTGTGAATCTGTTTGAACGCTGCG 1080  
DB 1208 AGCGTTCCCGGTACAGCGGATTAACGCGCGCGAGGTGTGAATCTGTTTGAACGCTGCG 1267  
QY 1081 CGTGGCGTACGATTAACGCGAGTGAATCAGTTTCTGCGTCTTCCAGACTTAAACAG 1140  
DB 1268 CGTGGCGTACGATTAACGCGAGTGAATCAGTTTCTGCGTCTTCCAGACTTAAACAG 1327  
QY 1141 CAGATGCGTATTAACGCGCGTGTCTTAATAACGCGCGCGAGAGTGAACCTGACG 1200  
DB 1328 CAGATGCGTATTAACGCGCGTGTCTTAATAACGCGCGCGAGAGTGAACCTGACG 1387  
QY 1201 CTGGCAGAGTGTGAAGAGCAAAATGCGCAGGCGATGTTGTTGCGTGGAGGTTTTCGCAA 1260  
DB 1388 CTGGCAGAGTGTGAAGAGCAAAATGCGCAGGCGATGTTGTTGCGTGGAGGTTTTCGCAA 1447  
QY 1261 ATGCTGAATGAAGCAGCATACCGGCTTGGCATTTTGA 1299  
DB 1448 ATGCTGAATGAAGCAGCATACCGGCTTGGCATTTTGA 1486

RESULT 7  
US-10-601-319-5  
Sequence 5, Application US/10601319  
Publication No. US20040091968A1  
GENERAL INFORMATION:  
APPLICANT: Short, Jay M.  
APPLICANT: Kretz, Keith A.  
APPLICANT: Gray, Kevin A.  
APPLICANT: Barton, Nelson Robert  
APPLICANT: Garrett, James B.  
APPLICANT: O' Donoghue, Eileen  
APPLICANT: Mathur, Eric J.  
TITLE OF INVENTION: RECOMBINANT PHYTASES AND METHODS OF MAKING  
TITLE OF INVENTION: AND USING THEM  
FILE REFERENCE: 09010-029011  
CURRENT APPLICATION NUMBER: US/10/601,319  
CURRENT FILING DATE: 2003-06-20  
PRIOR APPLICATION NUMBER: US 09/866,379  
PRIOR FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 09/580,515  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: US 09/318,528  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: US 09/291,931  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: US 09/259,214  
PRIOR FILING DATE: 1999-03-01

PRIOR APPLICATION NUMBER: US 08/910,798  
PRIOR FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 1901  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 403  
OTHER INFORMATION: n = A,T,C or G  
US-10-601-319-5

Query Match 99.3%; Score 1290; DB 17; Length 1901;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1293; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGAAAGCATCTTATCCCATTTTATCTCTGATTCCTGTAACCGGCAATCTGCA 60  
DB 188 ATGAAAGCATCTTATCCCATTTTATCTCTGATTCCTGTAACCGGCAATCTGCA 247  
QY 61 TTGCTCAGAGTGAGCGGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTGTCAATG 120  
DB 248 TTGCTCAGAGTGAGCGGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTGTCAATG 307  
QY 121 GTGCGTCTCCAAACCAAGGCGACGCACTGATGACAGAGATGTCACCCCAAGCATGCGCA 180  
DB 308 GTGCGTCTCCAAACCAAGGCGACGCACTGATGACAGAGATGTCACCCCAAGCATGCGCA 367  
QY 181 ACCTGGCCGGTAAACATGCGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240  
DB 368 ACCTGGCCGGTAAACATGCGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 427  
QY 241 GGAATTACCAAGCGGAGCTGTGTGAGCGAGATGTCGCGCAAAAAGGCTGCGCG 300  
DB 428 GGAATTACCAAGCGGAGCTGTGTGAGCGAGATGTCGCGCAAAAAGGCTGCGCGCG 487  
QY 301 CAGTGTGTCAAGTGTGCGATTAATGCTGATGTGACGAGCGTACCCGTAAACAGCGCA 360  
DB 488 CAGTGTGTCAAGTGTGCGATTAATGCTGATGTGACGAGCGTACCCGTAAACAGCGCA 547  
QY 361 GCGTTGGCGCGCGGCTGTGACCTGACCTGCAATTAACGTAATCCCAAGGAGATAGC 420  
DB 548 GCGTTGGCGCGCGGCTGTGACCTGACCTGCAATTAACGTAATCCCAAGGAGATAGC 607  
QY 421 TCCAGTCCGATCCGTTATTTAATCCCTTAAAACTGCGTTTCCCACTGATTAACGCG 480  
DB 608 TCCAGTCCGATCCGTTATTTAATCCCTTAAAACTGCGTTTCCCACTGATTAACGCG 667  
QY 481 AACGTGACGACGCGATCTCTCAGCAGGCGAGAGGCTCAATTGCTGACTTTACCGGCGAT 540  
DB 668 AACGTGACGACGCGATCTCTCAGCAGGCGAGAGGCTCAATTGCTGACTTTACCGGCGAT 727  
QY 541 CGGCAAAACGCGGTTTGGCGCAACTGGAACGGGTCTTAATTTTCCGCAATGAACCTGTGC 600  
DB 728 CGGCAAAACGCGGTTTGGCGCAACTGGAACGGGTCTTAATTTTCCGCAATGAACCTGTGC 787  
QY 601 CTTAAACGTGAGAAACAGACGAAGAGCTGTTCAATTAACGAGGCAATTAACGAGCACTC 660  
DB 788 CTTAAACGTGAGAAACAGACGAAGAGCTGTTCAATTAACGAGGCAATTAACGAGCACTC 847  
QY 661 AAGGTAGGCGCGACATGTCATTACCGGTCGTAAGCTCGCATCAATGCTGACG 720  
DB 848 AAGGTAGGCGCGACATGTCATTACCGGTCGTAAGCTCGCATCAATGCTGACG 907  
QY 721 GAGATATTTCTCTGCAACAAGCAAGGGAATGCGGAGCGGGGTGGGGGAAGATCACC 780  
DB 908 GAGATATTTCTCTGCAACAAGCAAGGGAATGCGGAGCGGGGTGGGGGAAGATCACC 967  
QY 781 GATTACACCAAGTGAACAACCTTGTATAGTTGCATTAACGCGAATTTTATTGCTACAA 840  
DB 968 GATTACACCAAGTGAACAACCTTGTATAGTTGCATTAACGCGAATTTTATTGCTACAA 1027

	Best Local Similarity	99.5%	Pred.No. 0;	Matches 1292;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
QY	1	ATGAAGCCGATCTTAA	TCCCATTTTTATCTCTTCTGATTTCCGTTAA	ACCCCGGCAATCTGCA	60			
Dp	188	ATGAAGCGATCTTAA	TATCCCATTTTTATCTCTTCTGATTTCCGTTAA	ACCCCGGCAATCTGCA	247			
QY	61	TTGCGCTCAGATGAGCGCGAGCTGAAAGCTGAAAGTGTGATTTGTCAGTGCATAGT	120					
Dp	248	TTGCTCTCAGATGAGCGCGAGCTGAAAGCTGAAAGTGTGATTTGTCAGTGCATAGT	307					
QY	121	GTGCGTCTCCAA	CCAAAGCCACGCAACTGATGCAAGGATGTCA	CCCCGAGCGATGCGCA	180			
Dp	308	GTGCGTCTCCAA	CCAAAGCCACGCAACTGATGCAAGGATGTCA	CCCCGAGCGATGCGCA	367			
QY	181	ACCTGCGCGGTAAAC	CTGGGTTGGCTGCAACACCGCGCGGTGTGAGCTAA	TGCGCTATCTC	240			
Dp	368	ACCTGCGCGGTAAAC	CTGGGTTGGCTGCAACACCGCGMGGTGTGAGCTAA	TGCGCTATCTC	427			
QY	241	GGA	CATTACCAAGCCACGCGCTGCTGTAACCGACGGAATGTGCTGGCGAAAAAGGCTGCGCG	300				
Dp	428	GGA	CATTACCTGGGCGACGCGTCTGTAACCGACGGAATGTGCTGGCGAAAAAGGCTGCGCGCG	487				
QY	301	CAGTCTGTGACG	GTGCGATTATGTGATGTGACGACGAGCGTAC	CCGTAAACAGCGGAA	360			
Dp	488	CAGTCTGTGACG	GTGCGATTATGTGATGTGACGACGAGCGTAC	CCGTAAACAGCGGAA	547			
QY	361	GCGTTGCGCGCGCGGCTGGGCACTGACTGTGCAATA	CCGTACTATCCACGACGAGATACG	420				
Dp	548	GCGTTGCGCGCGCGGCTGGGCACTGACTGTGCAATA	CCGTACTATCCACGACGAGAGATACG	607				
QY	421	TTCAGTCCGATCCG	TATTTAATCCCTPAAAAACGAGCGTTTGCAACTGATTAACGCG	480				

Qy	481	AAGGTACCTAACCGGATCTCCAGGAGGGACAGAGGGTCAATTGGTGACTTTACCGGGCAT	540
Db	668	AAGGTGACTACCGGATCTCCAGAGGGACAGAGGGTCAATTGGTGACTTTACCGGGCAT	727
Qy	541	CGGCAAAACGCGCTTTCCGCAACTGGAAACGGGGCTTTAATTTCCGCAATCAAACTTGTGC	600
Db	728	CGGCAAAACGCGCTTTCCGCAACTGGAAACGGGGCTTTAATTTCCGCAATCAAACTTGTGC	787
Qy	601	CTTAAACGTAGAAACAGACGAAAGCTTTCAATTACGACGATTAACATCGGAATC	660
Db	788	CTTAAACGTAGAAACAGACGAAAGCTTTCAATTACGACGATTAACATCGGAATC	847
Qy	661	AAGGTGAGCGCCGACAAATGTCTCAATTAAACGGGTGGGTAAAGCTTCGCATCATGTGACG	720
Db	848	AAGGTGAGCGCCGACAAATGTCTCAATTAAACGGGTGGGTAAAGCTTCGCATCATGTGACG	907
Qy	721	GAGTATTTTCTCTGCAACAAGCAACAGGAAATGCGGAGCCGGGTGGGGAAGATCAAC	780
Db	908	GAGTATTTTCTCTGCAACAAGCAACAGGAAATGCGGAGCCGGGTGGGGAAGATCAAC	967
Qy	781	GATTCAACCAAGTGGAAACACTTGTCTAAGTTTGATTAACGGCAATTTAATTTGGTACAA	840
Db	968	GATTCAACCAAGTGGAAACACTTGTCTAAGTTTGATTAACGGCAATTTAATTTGGTACAA	102
Qy	841	CGCACGCCAGAGGTTGGCCGCGAGCCGCGCAACCCCGTTATTGAATTGATCAAGACACG	900
Db	1028	CGCACGCCAGAGGTTGGCCGCGAGCCGCGCAACCCCGTTATTGAATTGATCAAGACACG	108
Qy	901	TTGACGCCCATCCACACCGCAAAAACAGGCGTATGTGTGACATTAACCACTTCAATGCTG	960
Db	1088	TTGACGCCCATCCACACCGCAAAAACAGGCGTATGTGTGACATTAACCACTTCAATGCTG	114
Qy	961	TTTATCGCCGGGACACGATACTAATCTGGAATCTCGCGCGCGGACCTGGAGCTCAACTGG	102
Db	1148	TTTATCGCCGGGACACGATACTAATCTGGAATCTCGCGCGCGGACCTGGAGCTCAACTGG	120
Qy	1021	ACCGTTCCCGGTACGCCGGAATTAACGCGCGCAGTGGTGAACGTGTGTAACGCTGG	1080



Db 1208 AGCGTCCCGGTACGCCGATTAACACGCCGACAGTGTGAATCTGTGTTTGAAGCTGG 1267  
Qy 1081 CGTCGGCTAAGCGATTAACGCCAGTGTGATTCAGTTCCTGTGCTTCCAGACTTTACG 1140  
Db 1268 CCGCGCTAAGCGATTAACGCCAGTGTGATTCAGTTCCTGTGCTTCCAGACTTTACG 1327  
Qy 1141 CAGATGCGGTAAACCGCGCTGTATTAATAGCCCGCCGGAAGGTGAACAGTAC 1200  
Db 1328 CAGATGCGGTAAACCGCGCTGTATTAATAGCCCGCCGGAAGGTGAACAGTAC 1387  
Qy 1201 CTGGCAGATGTGAAGCGAAATGCGCAGGCGATGTTCGTTGACAGTTTACGCA 1260  
Db 1388 CTGGCAGATGTGAAGCGAAATGCGCAGGCGATGTTCGTTGACAGTTTACGCA 1447  
Qy 1261 ATCGTGAATGAAGCAGCATACCGCTTGCACTTTGTA 1299  
Db 1448 ATCGTGAATGAAGCAGCATACCGCGCTGCACTTTGTA 1486  
RESULT 9  
US-10-601-319-6  
; Sequence 6, Application US/10601319  
; Publication No. US20040091968A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; APPLICANT: Kretz, Keith A.  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Barton, Nelson Robert  
; APPLICANT: Garrett, James B.  
; APPLICANT: O'Donoghue, Eileen  
; APPLICANT: Mathur, Eric J.  
; TITLE OF INVENTION: RECOMBINANT PHYTASES AND METHODS OF MAKING  
; FILE REFERENCE: 09010-029011  
; CURRENT APPLICATION NUMBER: US/10/601,319  
; PRIOR FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: US 09/866,379  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1901  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 403  
; OTHER INFORMATION: n = A,T,C or G  
US-10-601-319-6  
Query Match 99.2%; Score 1288.4; DB 17; Length 1901;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1292; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 308 GTGCGGTCCCAACCAAGGCCAGCACTGATGCAAGATGTACCCCAAGCCATGCGCA 367  
Qy 181 ACCGTGCGCGTAAACCTGGGTGGTGAACCGCGCGGTGGAGCTATCGCTATC 240  
Db 368 ACCGTGCGCGTAAACCTGGGTGGTGAACCGCGCGGTGGAGCTATCGCTATC 427  
Qy 241 GGAATTAACCAAGCGCAGCGTCTGGTGAAGCCAGATTCTGCGCAAAAAGGAGCTGCGC 300  
Db 428 GGAATTAACCAAGCGCAGCGTCTGGTGAAGCCAGATTCTGCGCAAAAAGGAGCTGCGC 487  
Qy 301 CAGTGTGTCAGATCGCATTTATGCTGATGCAAGACGTAACCCGTAAACAGCGAA 360  
Db 488 CAGTGTGTCAGATCGCATTTATGCTGATGCAAGACGTAACCCGTAAACAGCGAA 547  
Qy 361 GCGTTCGCGCGCGGCTGCGCACTGCTGCAATACCGTACATACCAAGGAGATAC 420  
Db 548 GCGTTCGCGCGCGGCTGCGCACTGCTGCAATACCGTACATACCAAGGAGATAC 607  
Qy 421 TCCAGTCCCGATCCGTTATTTAATCCTTAAACCTGGGCTTGGCACTGATACGCG 480  
Db 608 TCCAGTCCCGATCCGTTATTTAATCCTTAAACCTGGGCTTGGCACTGATACGCG 667  
Qy 481 AACGTACTGACGCGATCCTCAGCAGGCGAGAGGTCAATTGCTGACTTTACCGGCA 540  
Db 668 AACGTACTGACGCGATCCTCAGCAGGCGAGAGGTCAATTGCTGACTTTACCGGCA 727  
Qy 541 CGGCAACGCGCTTTCGCAACTGGAACGCGGTCTTAATTTTCGCAATCAACTGTG 600  
Db 728 CGGCAACGCGCTTTCGCAACTGGAACGCGGTCTTAATTTTCGCAATCAACTGTG 787  
Qy 601 CTAAACGTGAAGAAAGAGACGAAAGCTGTCAATTAACGAGGCAATTAACATGGAATC 660  
Db 788 CTAAACGTGAAGAAAGAGACGAAAGCTGTCAATTAACGAGGCAATTAACATGGAATC 847  
Qy 661 AAGGTAGCGCGCAATGCTCTTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 720  
Db 848 AAGGTAGCGCGCAATGCTCTTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 907  
Qy 721 GAGATATTTCTCTGCAACAGCAAGGAAATCCGAGCCGCGGTGGGGAAGATAC 780  
Db 908 GAGATATTTCTCTGCAACAGCAAGGAAATCCGAGCCGCGGTGGGGAAGATAC 967  
Qy 781 GATTCAACAGGTGAACCTTGTCTAAGTTGCTAATGCGCAATTTATTTGCTACAA 840  
Db 968 GATTCAACAGGTGAACCTTGTCTAAGTTGCTAATGCGCAATTTATTTGCTACAA 1027  
Qy 841 CGCAGCGCAGAGGTGCGCGCAGCGCCAGCCCGTTAATTGAATGATCAAGACGCG 900  
Db 1028 CGCAGCGCAGAGGTGCGCGCAGCGCCAGCCCGTTAATTGAATGATCAAGACGCG 1087  
Qy 901 TTGACGCCCATCATCACCGCAAAAGCGGTATGTGTGACATTAACCACTTACGTG 960  
Db 1088 TTGACGCCCATCATCACCGCAAAAGCGGTATGTGTGACATTAACCACTTACGTG 1147  
Qy 961 TTTATGCGCGGACAGATTAATCTGCAAAATCTGGCGCGCACTGAGGCTCAACTG 1020  
Db 1148 TTTATGCGCGGACAGATTAATCTGCAAAATCTGGCGCGCACTGAGGCTCAACTG 1207  
Qy 1021 ACGCTTCGCGTACGCGGATTAACGCGCGCAGGTGTGAATGCTGTTGAAGCTG 1080  
Db 1208 ACGCTTCGCGTACGCGGATTAACGCGCGCAGGTGTGAATGCTGTTGAAGCTG 1267  
Qy 1081 CGTCGGCTAAGCGATTAACGCCAGTGTGATTCAGTTCCTGTGCTTCCAGACTTTACG 1140  
Db 1268 CGTCGGCTAAGCGATTAACGCCAGTGTGATTCAGTTCCTGTGCTTCCAGACTTTACG 1327  
Qy 1141 CAGATGCGGTAAACCGCGCTGTATTAATAGCCCGCCGGAAGGTGAACAGTAC 1200  
Db 1328 CAGATGCGGTAAACCGCGCTGTATTAATAGCCCGCCGGAAGGTGAACAGTAC 1387  
Qy 1201 CTGGCAGATGTGAAGCGAAATGCGCAGGCGATGTTCGTTGACAGTTTACGCA 1260

Db 1388 CTGGCAGGATGTGAAGAGGAATAGCGAGGCATGTGTGTCGACAGTTTACGAA 1447  
Qy 1261 ATCGTAATGAAGCAGCATACCCGCTTGCAAGTTTGTA 1299  
Db 1448 ATCGTAATGAAGCAGCATACCCGCTTGCAAGTTTGTA 1486

RESULT 10  
US-09-777-566A-1  
; Sequence 1, Application US/09777566A  
; Patent No. US20010055788A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay  
; APPLICANT: KRETZ, Keith  
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
; FILE REFERENCE: DIVER1370-6  
; CURRENT APPLICATION NUMBER: US/09/777,566A  
; PRIOR FILING DATE: 2001-06-11  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1320)  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1323)  
; OTHER INFORMATION: n is any nucleotide  
US-09-777-566A-1

Query Match 98.7%; Score 1282.2; DB 9; Length 1323;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGATTCGTTAAACCCCGCAATCTGA 60  
Db 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGATTCGTTAAACCCCGCAATCTGA 60  
Qy 61 TTGCGTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTCTCAAGT 120  
Db 61 TTGCGTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTCTCAAGT 120  
Qy 121 GTGCGTGTCTCAACCAAGGCAAGCACTGATGCAAGATGTCACCCCAAGCGATGGCA 180  
Db 121 GTGCGTGTCTCAACCAAGGCAAGCACTGATGCAAGATGTCACCCCAAGCGATGGCA 180  
Qy 121 GTGCGTGTCTCAACCAAGGCAAGCACTGATGCAAGATGTCACCCCAAGCGATGGCA 180  
Db 121 GTGCGTGTCTCAACCAAGGCAAGCACTGATGCAAGATGTCACCCCAAGCGATGGCA 180  
Qy 181 ACCTGACCGGTAAACTGGGTTGGTGAACAACCGCGGTGTGAGCTAATCGCTATCTC 240  
Db 181 ACCTGACCGGTAAACTGGGTTGGTGAACAACCGCGGTGTGAGCTAATCGCTATCTC 240  
Qy 241 GGCATTAACCAAGCCAGCGCTCTGTGTAAGCCGAGTGTGCGGAAAAAGGGCTGCCG 300  
Db 241 GGCATTAACCAAGCCAGCGCTCTGTGTAAGCCGAGTGTGCGGAAAAAGGGCTGCCG 300  
Qy 301 CAGCTGTGTCAGGTCCGATTTATGTCTGATGTCAGAGCGTACCCGTAACAGGCGAA 360  
Db 301 CAGCTGTGTCAGGTCCGATTTATGTCTGATGTCAGAGCGTACCCGTAACAGGCGAA 360  
Qy 361 GCGTTGCGCGCGCGGCTGCGCACTGACTGTGCAATTAACCTGACATACCCGAGGAGATAG 420  
Db 361 GCGTTGCGCGCGCGGCTGCGCACTGACTGTGCAATTAACCTGACATACCCGAGGAGATAG 420

Qy 421 TCCAGTCCCATCCGTTATTTAATCTCTAATAAACTGCGGTTTGCACAACTGGAATACGG 480  
Db 421 TCCAGTCCCATCCGTTATTTAATCTCTAATAAACTGCGGTTTGCACAACTGGAATACGG 480  
Qy 481 AACGTACTGAACCGATCTCTAGCAGGCGAGAGGTCATTTGTCATTTACCGGGCAT 540  
Db 481 AACGTACTGAACCGATCTCTAGCAGGCGAGAGGTCATTTGTCATTTACCGGGCAT 540  
Qy 541 CCGGCAACCGGCTTTCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAACTTGTGC 600  
Db 541 CCGGCAACCGGCTTTCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAACTTGTGC 600  
Qy 601 CTTAAACGTGAAGAACAGAGCAAGAGCTGTTCAATTAACGAGGATTAACATCGAACTC 660  
Db 601 CTTAAACGTGAAGAACAGAGCAAGAGCTGTTCAATTAACGAGGATTAACATCGAACTC 660  
Qy 661 AAGGTAGCGCGCAATGTCTCATTTAACCGGTGCGTGAAGCTTCGATCAATGTCAGC 720  
Db 661 AAGGTAGCGCGCAATGTCTCATTTAACCGGTGCGTGAAGCTTCGATCAATGTCAGC 720  
Qy 721 GAGATATTTCTCTGCAACAGCAAGGGAATGCGAGCGGGGTGGGAAAGATCAC 780  
Db 721 GAGATATTTCTCTGCAACAGCAAGGGAATGCGAGCGGGGTGGGAAAGATCAC 780  
Qy 781 GATTCAACACAGTGAACACCTTGCTAAGTTTGCAATACGCGCAATTTATTTGCTACAA 840  
Db 781 GATTCAACACAGTGAACACCTTGCTAAGTTTGCAATACGCGCAATTTATTTGCTACAA 840  
Qy 841 CGCACGCAAGAGTTGCGCGCAAGCGCGCACCCCGTTATTTAGATTGATCATGACAGC 900  
Db 841 CGCACGCAAGAGTTGCGCGCAAGCGCGCACCCCGTTATTTAGATTGATCATGACAGC 900  
Qy 901 TTGACGCCCATTCACCGCAAAAACAGCGATGATGTCATTAACCATTCAGTGTCTG 960  
Db 901 TTGACGCCCATTCACCGCAAAAACAGCGATGATGTCATTAACCATTCAGTGTCTG 960  
Qy 961 TTTATTCGCGGACAGATACTAATCTGSCAAATCTCGCGGCGCACTGAGCTCACTGG 1020  
Db 961 TTTATTCGCGGACAGATACTAATCTGSCAAATCTCGCGGCGCACTGAGCTCACTGG 1020  
Qy 1021 ACGTTCCCGGTCAAGCCGATTAACCGCGGCGATGATGTCATGATTTGAAAGCTGG 1080  
Db 1021 ACGTTCCCGGTCAAGCCGATTAACCGCGGCGATGATGTCATGATTTGAAAGCTGG 1080  
Qy 1081 CGTGGCTAAGGATTAACAGCAAGTGAATTCAGGTTTGCTGCTTCCAGACTTTACAG 1140  
Db 1081 CGTGGCTAAGGATTAACAGCAAGTGAATTCAGGTTTGCTGCTTCCAGACTTTACAG 1140  
Qy 1141 CAGATGCGGTAAACCGCGCTGTCAATTAATACCGCGCGGAGAGTGAACCTGACC 1200  
Db 1141 CAGATGCGGTAAACCGCGCTGTCAATTAATACCGCGCGGAGAGTGAACCTGACC 1200  
Qy 1201 CTGGCAGATGTGAAGAGCGAAATCGCAGGCGATGTGTTGTTGCGAGTTTACGAA 1260  
Db 1201 CTGGCAGATGTGAAGAGCGAAATCGCAGGCGATGTGTTGTTGCGAGTTTACGAA 1260  
Qy 1261 ATCGTAATGAAGCAGCATACCCGCTTGCAAGTTTG 1296  
Db 1261 ATCGTAATGAAGCAGCATACCCGCTTGCAAGTTTG 1296

RESULT 11  
US-09-866-379-1  
; Sequence 1, Application US/09866379  
; Patent No. US20020136754A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay  
; APPLICANT: KRETZ, Keith  
; APPLICANT: GRAY, Kevin  
; APPLICANT: BARTON, Nelson  
; APPLICANT: GARRETT, James  
; APPLICANT: O'DONOGHUE, Eileen

```

; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1323)
; OTHER INFORMATION: n is any nucleotide
; NAME/KEY: CDS
; LOCATION: (1)..(1323)
; OTHER INFORMATION:
; US-09-866-379-1

```

```

Query Match      98.7%; Score 1282.2; DB 9; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 ATGAAGCCATCTTAATCCATTTTATCTCTCTGATTCGGTAAACCCGCAATCTGCA 60
DB 1 ATGAAGCCATCTTAATCCATTTTATCTCTCTGATTCGGTAAACCCGCAATCTGCA 60
QY 61 TTGCGCTCAGAGTAGAGCCGAGCTGAAGCTGAAAGTGTGATGTCACTGCTCATGCT 120
DB 61 TTGCGCTCAGAGTAGAGCCGAGCTGAAGCTGAAAGTGTGATGTCACTGCTCATGCT 120
QY 121 GTTCGCTCAGAGTAGAGCCGAGCTGAAGCTGAAAGTGTGATGTCACTGCTCATGCT 180
DB 121 GTTCGCTCAGAGTAGAGCCGAGCTGAAGCTGAAAGTGTGATGTCACTGCTCATGCT 180
QY 181 ACCTGCCCGGTAAACCTGGGTTGAGCTGACACCGCGCGGTGAGTAACTGCTATCTC 240
DB 181 ACCTGCCCGGTAAACCTGGGTTGAGCTGACACCGCGCGGTGAGTAACTGCTATCTC 240
QY 241 GACATTTACCAACGCCAGCGTCTGTGATGACCGAGCGGATTCCTGCGGAAAAAGGCTGCCG 300
DB 241 GACATTTACCAACGCCAGCGTCTGTGATGACCGAGCGGATTCCTGCGGAAAAAGGCTGCCG 300
QY 301 CAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 CAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 GCGCTGCGCGCGCGCGCGCGCGCTGACCTGATGATGATGATGATGATGATGATGATGAT 420
DB 361 GCGCTGCGCGCGCGCGCGCGCGCTGACCTGATGATGATGATGATGATGATGATGATGAT 420
QY 421 TCCAGTCCGATCCGATATTTAATCTCTAATAAACTGCGCTTTCGCACTGATTAACGCG 480
DB 421 TCCAGTCCGATCCGATATTTAATCTCTAATAAACTGCGCTTTCGCACTGATTAACGCG 480
QY 481 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 481 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 CCGGCAACCGGCGGTTTCGCGAACTGGAACCGGCTGCTAATTTTCGCGAATCAAACTTGTGC 600
DB 541 CCGGCAACCGGCGGTTTCGCGAACTGGAACCGGCTGCTAATTTTCGCGAATCAAACTTGTGC 600

```

```

QY 601 CTTAAACGTGAGAAACAGGACGAAGCTGTCATTAAACGACGATTAACATCGAATCTC 660
DB 601 CTTAAACGTGAGAAACAGGACGAAGCTGTCATTAAACGACGATTAACATCGAATCTC 660
QY 661 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 661 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 GAGATATTTCTCTCTGCAACAGCAACGGAATGCGGACCGCGGCGGGAAGATCACG 780
DB 721 GAGATATTTCTCTCTGCAACAGCAACGGAATGCGGACCGCGGCGGGAAGATCACG 780
QY 781 GATTACACAGTGAACACCTTGCTAATGTTGATTAACGCGAATTTATTTGCTACAA 840
DB 781 GATTACACAGTGAACACCTTGCTAATGTTGATTAACGCGAATTTATTTGCTACAA 840
QY 841 CGCAGCCAGAGGTTGCGCGAGCCGCGCACCCCGTTATTAAGTTGATCAAGACACG 900
DB 841 CGCAGCCAGAGGTTGCGCGAGCCGCGCACCCCGTTATTAAGTTGATCAAGACACG 900
QY 901 TTGACCGCCCATTCACCGCAAAACAGGCGTATGATGATGATGATGATGATGATGATGAT 960
DB 901 TTGACCGCCCATTCACCGCAAAACAGGCGTATGATGATGATGATGATGATGATGATGAT 960
QY 961 TTTATGCGCGGACAGATCTAATCTGCAAAATCTGCGCGGCGCACTGAGCTCAACTGG 1020
DB 961 TTTATGCGCGGACAGATCTAATCTGCAAAATCTGCGCGGCGCACTGAGCTCAACTGG 1020
QY 1021 ACAGCTTCCCGGTGACCGGATTAACAGCGCGCGAGTGTGAATCTGATGTTGAACCTGG 1080
DB 1021 ACAGCTTCCCGGTGACCGGATTAACAGCGCGCGAGTGTGAATCTGATGTTGAACCTGG 1080
QY 1081 CGTGGCTAAGCGATTAACAGCGCGAGTGTGAATCTGATGTTGAACCTGG 1140
DB 1081 CGTGGCTAAGCGATTAACAGCGCGAGTGTGAATCTGATGTTGAACCTGG 1140
QY 1141 CAGATCGGTGATTAACAGCGCGAGTGTGAATCTGATGTTGAACCTGG 1200
DB 1141 CAGATCGGTGATTAACAGCGCGAGTGTGAATCTGATGTTGAACCTGG 1200
QY 1201 CTGCGAGATGATGAAGAGCAAAATGCGCAGGCGATGTTGTTGCGAGGTTTTCGCA 1260
DB 1201 CTGCGAGATGATGAAGAGCAAAATGCGCAGGCGATGTTGTTGCGAGGTTTTCGCA 1260
QY 1261 ATGCTGAATGAACAGCATATCCGCTTGCACTTTG 1296
DB 1261 ATGCTGAATGAACAGCATATCCGCTTGCACTTTG 1296

```

```

RESULT 12
US-10-034-985-1
; Sequence 1, Application US/10034985
; Publication No. US20030049815A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/034,985
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-10-034-985-1

Query Match      98.7%; Score 1282.2; DB 14; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGGTTAAACCCGCAATCTGCA 60
DB 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGGTTAAACCCGCAATCTGCA 60
QY 61 TTGCGTCAGAGTGAAGCCGAGCTGAAAGTGAAGTGTGATTTGTCAGTGTATGAT 120
DB 61 TTGCGTCAGAGTGAAGCCGAGCTGAAAGTGAAGTGTGATTTGTCAGTGTATGAT 120
QY 121 GTGCGTGTCCAAACCAAGCCCAACGCACTGATGCAAGATGTACCCCAACGCAATG 180
DB 121 GTGCGTGTCCAAACCAAGCCCAACGCACTGATGCAAGATGTACCCCAACGCAATG 180
QY 121 GTGCGTGTCCAAACCAAGCCCAACGCACTGATGCAAGATGTACCCCAACGCAATG 180
DB 121 GTGCGTGTCCAAACCAAGCCCAACGCACTGATGCAAGATGTACCCCAACGCAATG 180
QY 181 ACCTGCGCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240
DB 181 ACCTGCGCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240
QY 241 GGACATTAACCAACGCCAGCGCTGTGAGCCGACGAAATTCGTGGCGAAAAAGGCGTGGCCG 300
DB 241 GGACATTAACCAACGCCAGCGCTGTGAGCCGACGAAATTCGTGGCGAAAAAGGCGTGGCCG 300
QY 301 CAGTGTGTGACAGTGTCCGCAATTAATGCTGATGTCAGACGACCTAACCCGTTAAACAGGCCGA 360
DB 301 CAGTGTGTGACAGTGTCCGCAATTAATGCTGATGTCAGACGACCTAACCCGTTAAACAGGCCGA 360
QY 361 GCCTTGCGCGCGCGGCTGGCACTGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 420
DB 361 GCCTTGCGCGCGCGGCTGGCACTGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 420
QY 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCGTTTGGCACTGATTAACGCG 480
DB 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCGTTTGGCACTGATTAACGCG 480
QY 481 AACGTACTGAACCCGCAATCTCTGACAGGCAAGGCTCAATGCTCAATGCTCAATGCTCAATG 540
DB 481 AACGTACTGAACCCGCAATCTCTGACAGGCAAGGCTCAATGCTCAATGCTCAATGCTCAATG 540
QY 541 CGGCAAAACGCGGTTTGGCGAACTGAAACGCGGCTTAAATTTTCCGCAATCAAACTTGTC 600
DB 541 CGGCAAAACGCGGTTTGGCGAACTGAAACGCGGCTTAAATTTTCCGCAATCAAACTTGTC 600
QY 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTTACCATGGAATCTC 660
DB 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTTACCATGGAATCTC 660
QY 661 AAGGTAGCGCCGCAATGCTCAATTAACGAGCATTTACCATGGAATCTC 720
DB 661 AAGGTAGCGCCGCAATGCTCAATTAACGAGCATTTACCATGGAATCTC 720
QY 721 GAGATATTTCTCTGCAACAGACAGGAAATCCCGAGCCGCGGTTGGGGAAGATCAACC 780
DB 721 GAGATATTTCTCTGCAACAGACAGGAAATCCCGAGCCGCGGTTGGGGAAGATCAACC 780
QY 781 GATTCAACACCAAGTGAACACTTGTAAAGTTTGCATTAACGCAATTTTATTGTTACAA 840
DB 781 GATTCAACACCAAGTGAACACTTGTAAAGTTTGCATTAACGCAATTTTATTGTTACAA 840
QY 841 CGACAGCCAGAGGTGCGCGAGCCGCGCAACCCCGTTAATTAATTTGATCAAGACAGG 900
DB 841 CGACAGCCAGAGGTGCGCGAGCCGCGCAACCCCGTTAATTAATTTGATCAAGACAGG 900
QY 901 TTGAGCGCCCATCAACGCAAAACAGGCGTATGTGATTAATCACTTCACTGATGCTG 960

```

```

DB 901 TTGAGCGCCCATCAACGCAAAACAGGCGTATGTGATTAATCACTTCACTGATGCTG 960
QY 961 TTATATCGCGGACACGATACTTAATCTGCAAAATCTCGCGCGGCACTGAGCTCAACTG 1020
DB 961 TTATATCGCGGACACGATACTTAATCTGCAAAATCTCGCGCGGCACTGAGCTCAACTG 1020
QY 1021 ACGTTCGCCGCTCAACCGGATTAACAGCCGCGCAGAGTGTGAATCTGTGTTGAACGCTG 1080
DB 1021 ACGTTCGCCGCTCAACCGGATTAACAGCCGCGCAGAGTGTGAATCTGTGTTGAACGCTG 1080
QY 1081 CGTGGCTTAACCGATTAACAGCCAGTGAATTCAGTTTCGCTGTGCTTTCAGACTTTACAG 1140
DB 1081 CGTGGCTTAACCGATTAACAGCCAGTGAATTCAGTTTCGCTGTGCTTTCAGACTTTACAG 1140
QY 1141 CAGATGCGTGAATTAACCCGCTGTCAATTAATACGCCCGCGGAGAGGTGAACCTGACC 1200
DB 1141 CAGATGCGTGAATTAACCCGCTGTCAATTAATACGCCCGCGGAGAGGTGAACCTGACC 1200
QY 1201 CTGGCAGAGTGAAGACGAAATGCGAGGCAATGTTGCTTGGCAGTGTTTAACGCA 1260
DB 1201 CTGGCAGAGTGAAGACGAAATGCGAGGCAATGTTGCTTGGCAGTGTTTAACGCA 1260
QY 1261 ATCGTAATGAAGACGCAATACCGCTTGCAATTG 1296
DB 1261 ATCGTAATGAAGACGCAATACCGCTTGCAATTG 1296

```

```

RESULT 13
US-10-430-356-1
; Sequence 1, Application US/10430356
; Publication No. US20030232041A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/430,356
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-10-430-356-1

```

```

Query Match      98.7%; Score 1282.2; DB 17; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGGTTAAACCCGCAATCTGCA 60
DB 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGGTTAAACCCGCAATCTGCA 60
QY 61 TTGCGTCAGAGTGAAGCCGAGCTGAAAGTGAAGTGTGATTTGTCAGTGTATGAT 120

```

Db 61 TTCGCTCAGATGAGCCGAGCTGAAAGCTGGAAGATGTGATGATTTGTCAATGCT 120  
Qy 121 GTGCGTGTCTCAACCAAGGCCACGCACTGATGAGAGATGTGACCCGACGATGCGCA 180  
Db 121 GTGCGTGTCTCAACCAAGGCCACGCACTGATGAGAGATGTGACCCGACGATGCGCA 180  
Qy 181 ACCTGGCCGGTAAACTGGGTGGCTGACACCGCGGGTGGTGAAGTAAATCCCTATCTC 240  
Db 181 ACCTGGCCGGTAAACTGGGTGGCTGACACCGCGGGTGGTGAAGTAAATCCCTATCTC 240  
Qy 241 GGAACATTAACCAAGCCAGCGCTGAGTGAAGCGATGCTGGCGCAAAAAGGCTGCGCG 300  
Db 241 GGAACATTAACCAAGCCAGCGCTGAGTGAAGCGATGCTGGCGCAAAAAGGCTGCGCG 300  
Qy 301 CAGCTGTGTCAGAGTCGCGATTAATGCTGATGTCAGAGCGTACCCGTAACAGGCGAA 360  
Db 301 CAGCTGTGTCAGAGTCGCGATTAATGCTGATGTCAGAGCGTACCCGTAACAGGCGAA 360  
Qy 361 GCGTTCGCGCGCGGGCTGGCACTGACTGTGCAATACCGTACATACCCAGGCAATAG 420  
Db 361 GCGTTCGCGCGCGGGCTGGCACTGACTGTGCAATACCGTACATACCCAGGCAATAG 420  
Qy 421 TCCAGTCCCGATCCGTTATTAATCTCTAATAAACTGGGTTGCCAATGATTAACGCG 480  
Db 421 TCCAGTCCCGATCCGTTATTAATCTCTAATAAACTGGGTTGCCAATGATTAACGCG 480  
Qy 481 AACGTACTGACGCGATCTCTACAGCAGGAGGAGGATGCAATGCTGATCCGAGCAT 540  
Db 481 AACGTACTGACGCGATCTCTACAGCAGGAGGAGGATGCAATGCTGATCCGAGCAT 540  
Qy 541 CCGCAACAGCGCGTGGCGCACTGGAACGGGTCTTAATTTTCCGCAATCAACTGTGCG 600  
Db 541 CCGCAACAGCGCGTGGCGCACTGGAACGGGTCTTAATTTTCCGCAATCAACTGTGCG 600  
Qy 601 CTAAACGTGAAGAAAGAGAGCGAAGCTGTTCTTAATACGAGGCAATTAACATGAGAACTC 660  
Db 601 CTAAACGTGAAGAAAGAGAGCGAAGCTGTTCTTAATACGAGGCAATTAACATGAGAACTC 660  
Qy 661 AAGGTAGGCGCGCAATGCTCTATTAAACGGTGGGTAAGCCTCGATCAATGCTGAGC 720  
Db 661 AAGGTAGGCGCGCAATGCTCTATTAAACGGTGGGTAAGCCTCGATCAATGCTGAGC 720  
Qy 721 GAGATATTTCTCTCTGCAACAGCAACAGGAATGCCGAGCGGGGTGGGGAAGATCAC 780  
Db 721 GAGATATTTCTCTCTGCAACAGCAACAGGAATGCCGAGCGGGGTGGGGAAGATCAC 780  
Qy 781 GATTCAACACAGTGAACAACCTGTGTAAGTTGATTAACGCGCAATTTATTTGCTACAA 840  
Db 781 GATTCAACACAGTGAACAACCTGTGTAAGTTGATTAACGCGCAATTTATTTGCTACAA 840  
Qy 841 CGCAGCGCAGAGGTTGGCCGACCGCGCCACCCCGTTAATGATTTGATCAAGACAGCG 900  
Db 841 CGCAGCGCAGAGGTTGGCCGACCGCGCCACCCCGTTAATGATTTGATCAAGACAGCG 900  
Qy 901 TTGACGCGCCATTCACCGCAAAAACAGGCGTATGTGTGACATTAACCATTCAGTACTG 960  
Db 901 TTGACGCGCCATTCACCGCAAAAACAGGCGTATGTGTGACATTAACCATTCAGTACTG 960  
Qy 961 TTTATGCGCGGAGACAGTATTAATCTGAGCAATCTGGCGCGCATGAGCTCAATGCG 1020  
Db 961 TTTATGCGCGGAGACAGTATTAATCTGAGCAATCTGGCGCGCATGAGCTCAATGCG 1020  
Qy 1021 ACCGTTCCCGATCAGCGGATTAACAGCGCGCAGAGTGTGAATCTGTGTTGAACGCTGG 1080  
Db 1021 ACCGTTCCCGATCAGCGGATTAACAGCGCGCAGAGTGTGAATCTGTGTTGAACGCTGG 1080  
Qy 1081 CGTGGCTTAAGCATTAACAGCAGTGAATTCAGGTTTCTGCTGCTTCAGACTTTTACAG 1140  
Db 1081 CGTGGCTTAAGCATTAACAGCAGTGAATTCAGGTTTCTGCTGCTTCAGACTTTTACAG 1140  
Qy 1141 CAGATGGGTATTAAGAGCGCGCTGTCTTAATATGCGCGCCGAGAGGTGAACCTGACC 1200  
Db 1141 CAGATGGGTATTAAGAGCGCGCTGTCTTAATATGCGCGCCGAGAGGTGAACCTGACC 1200

Qy 1201 CTGCGAGATGAGAGAGCAAAATGCGAGGCGATGTGCTGTGGCAGGTTTACGCA 1260  
Db 1201 CTGCGAGATGAGAGAGCAAAATGCGAGGCGATGTGCTGTGGCAGGTTTACGCA 1260  
Qy 1261 ATCGTAATGAAGCAGCATACCGCTTGACATTTG 1296  
Db 1261 ATCGTAATGAAGCAGCATACCGCGTGCAGATTTG 1296  
RESULT 14  
US-10-601-319-1  
; Sequence 1, Application US/10601319  
; Publication No. US20040091968A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; APPLICANT: Kretz, Keith A.  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Barton, Nelson Robert  
; APPLICANT: Garrett, James B.  
; APPLICANT: O'Donoghue, Eileen  
; APPLICANT: Mathur, Eric J.  
; TITLE OF INVENTION: RECOMBINANT PHYLASES AND METHODS OF MAKING  
; FILE REFERENCE: 09010-029011  
; CURRENT APPLICATION NUMBER: US/10/601,319  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: US 09/866,379  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1320)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 216  
; OTHER INFORMATION: n = A,T,C or G  
US-10-601-319-1  
Query Match 98.7%; Score 1282.2; DB 17; Length 1323;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 ATGAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTTAACCCGCAATCTGCA 60  
Db 1 ATGAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTTAACCCGCAATCTGCA 60  
Qy 61 TTGCTCAGAGTGAAGCCGAGCTGGAAGCTGGAAAGTGTGTGATTTGTCATTCGTCATG 120  
Db 61 TTGCTCAGAGTGAAGCCGAGCTGGAAGCTGGAAAGTGTGTGATTTGTCATTCGTCATG 120  
Qy 121 GTGCGTGTCTCAACCAAGGCCACGCACTGATGAGAGATGTGACCCGACGATGCGCA 180  
Db 121 GTGCGTGTCTCAACCAAGGCCACGCACTGATGAGAGATGTGACCCGACGATGCGCA 180  
Qy 181 ACCTGGCCGGTAAACTGGGTGGCTGACACCGCGGGTGGTGAAGTAAATCCCTATCTC 240  
Db 181 ACCTGGCCGGTAAACTGGGTGGCTGACACCGCGGGTGGTGAAGTAAATCCCTATCTC 240

```

QY 241 GGAATTATACCAACGCGAGCTGTGTAGCCGACGGAATTGCTGGGAAAAGGGCGTCCCG 300
DB 241 GGAATTATACCAACGCGAGCTGTGTAGCCGACGGAATTGCTGGGAAAAGGGCGTCCCG 300
QY 301 CAGTCTGTGAGGTGCGGATTTATGCTGTGACGACGCTGACCCGTAACAGCGCA 360
DB 301 CAGTCTGTGAGGTGCGGATTTATGCTGTGACGACGCTGACCCGTAACAGCGCA 360
QY 361 GCTTTCGCGCGCGGCTGGACCTGACCTGTGCAATTAACCTGATACCCAGCAATACG 420
DB 361 GCTTTCGCGCGCGGCTGGACCTGACCTGTGCAATTAACCTGATACCCAGCAATACG 420
QY 421 TCCAGTCCCGATCCGTTATTTATTAATCTCTAAACCTGGGCTTGGCACTGATTAACGG 480
DB 421 TCCAGTCCCGATCCGTTATTTATTAATCTCTAAACCTGGGCTTGGCACTGATTAACGG 480
QY 481 AACGTGACTGACGCGATCTCTAGACGAGGAGGATCAATTGCTGACTTTACCGGGCAT 540
DB 481 AACGTGACTGACGCGATCTCTAGACGAGGAGGATCAATTGCTGACTTTACCGGGCAT 540
QY 541 CGGCAAAACGCGCTTTCGCGAATCTGGAACGCGGTCTTAATTTCCGCAATCAACTTGTGC 600
DB 541 CGGCAAAACGCGCTTTCGCGAATCTGGAACGCGGTCTTAATTTCCGCAATCAACTTGTGC 600
QY 601 CTTAAACGTAAGAAACAGGACGAAAGCTGTTCATTAAACGAGGATTAACATCGGAATC 660
DB 601 CTTAAACGTAAGAAACAGGACGAAAGCTGTTCATTAAACGAGGATTAACATCGGAATC 660
QY 661 AAGGTAGCGCGGACCAATGTCATTAAACGCGGTGCTTAAGCTCTGCAATGCTGACG 720
DB 661 AAGGTAGCGCGGACCAATGTCATTAAACGCGGTGCTTAAGCTCTGCAATGCTGACG 720
QY 721 GAGATTTTCTCTGTGAACAGACAGGAAATCCCGAGCCGCGGGGGGAAAGATCAAC 780
DB 721 GAGATTTTCTCTGTGAACAGACAGGAAATCCCGAGCCGCGGGGGGAAAGATCAAC 780
QY 781 GATTACACACAGTGAACACCTTGTGATTTGATTAACGCGCAATTTATTTGCTACAA 840
DB 781 GATTACACACAGTGAACACCTTGTGATTTGATTAACGCGCAATTTATTTGCTACAA 840
QY 841 CGCAGCCGACGAGGTTGCGCGACGCGGCAACCCGTTATTAAGATTGATCAAGACAGCG 900
DB 841 CGCAGCCGACGAGGTTGCGCGACGCGGCAACCCGTTATTAAGATTGATCAAGACAGCG 900
QY 901 TTGACGCGCCCATCTCACCGCAAAAACAGGCTATGCTGACATTAACCTCAATGATCG 960
DB 901 TTGACGCGCCCATCTCACCGCAAAAACAGGCTATGCTGACATTAACCTCAATGATCG 960
QY 961 TTTATCGCGGACACGATTAATCTGCAAAATCTGCGGCGGCACTGAGCTCAACTCG 1020
DB 961 TTTATCGCGGACACGATTAATCTGCAAAATCTGCGGCGGCACTGAGCTCAACTCG 1020
QY 1021 ACGCTTCCCGGTACGCGGATTAACGCGCGGAGTGTGAATGCTGTTGAACCTCG 1080
DB 1021 ACGCTTCCCGGTACGCGGATTAACGCGCGGAGTGTGAATGCTGTTGAACCTCG 1080
QY 1081 ACGCTTCCCGGTACGCGGATTAACGCGCGGAGTGTGAATGCTGTTGAACCTCG 1140
DB 1081 ACGCTTCCCGGTACGCGGATTAACGCGCGGAGTGTGAATGCTGTTGAACCTCG 1140
QY 1141 CAGATCGGTGAATAAAGCGCGCTGTCAATTAATACGCGCGGAGAGGTGAATGAC 1200
DB 1141 CAGATCGGTGAATAAAGCGCGCTGTCAATTAATACGCGCGGAGAGGTGAATGAC 1200
QY 1201 CTGGCAGAGATGAAGAAGCGAAATGCGCAGGCGCATGCTTGTGGAGGTTTACGCA 1260
DB 1201 CTGGCAGAGATGAAGAAGCGAAATGCGCAGGCGCATGCTTGTGGAGGTTTACGCA 1260
QY 1261 ATGTGAATGAACGACGATTAACCGGCTGAGTTG 1296
DB 1261 ATGTGAATGAACGACGATTAACCGGCTGAGTTG 1296

```

```

RESULT 15
US-10-156-660-1
; Sequence 1, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Bileen
; APPLICANT: Machuga, Eric J.
; TITLE OF INVENTION: PHYLASES, NUCLEIC ACIDS ENCODING THEM
; FILE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: modified phylase enzyme
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-10-156-660-1

Query Match          96.9%; Score 1259.2; DB 15; Length 1308;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 ATGAAAGCATTTTATCCATTTTATCTCTTGATTCGTTAAACCCGCAATCTGCA 60
DB 1 ATGAAAGCATTTTATCCATTTTATCTCTTGATTCGTTAAACCCGCAATCTGCA 60
QY 61 TTGCGTCAGAGTACCGGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTGTGATG 120
DB 61 TTGCGTCAGAGTACCGGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTGTGATG 120
QY 121 GTGCGTCTCCAAACAGGCGCAACGCACTGATGACGATGTCACCCAGACGATGGCA 180
DB 121 GTGCGTCTCCAAACAGGCGCAACGCACTGATGACGATGTCACCCAGACGATGGCA 180
QY 121 GTGCGTCTCCAAACAGGCGCAACGCACTGATGACGATGTCACCCAGACGATGGCA 180
DB 121 GTGCGTCTCCAAACAGGCGCAACGCACTGATGACGATGTCACCCAGACGATGGCA 180
QY 181 ACTGCGCGGTAAACTGGGTGCTGACACCGCGGGGTGAGCTAATGCGCTATCTC 240
DB 181 ACTGCGCGGTAAACTGGGTGCTGACACCGCGGGGTGAGCTAATGCGCTATCTC 240
QY 241 GGAATTATACCAACGCGAGCTGTGTAACGCGATTTGTCGCAAAAAGGGCTGCCG 300
DB 241 GGAATTATACCAACGCGAGCTGTGTAACGCGATTTGTCGCAAAAAGGGCTGCCG 300
QY 301 CAGTCTGTGAGGTGCGGATTTATGCTGTGACGACGCTGACCCGTAACAGCGCA 360
DB 301 CAGTCTGTGAGGTGCGGATTTATGCTGTGACGACGCTGACCCGTAACAGCGCA 360
QY 361 GCTTTCGCGCGCGGCTGGACCTGACCTGTGCAATTAACCTGATACCCAGCAATACG 420
DB 361 GCTTTCGCGCGCGGCTGGACCTGACCTGTGCAATTAACCTGATACCCAGCAATACG 420

```

Db 361 GCCTTGCCGCGGCGTGGCACCTGACTGCAATACCGTACATACCAGGCAATACG 420  
Qy 421 TCCAGTCCCGATCCGTTATTTAACTCTTAAAACTGCGTTTGCCACTGATTAACGCG 480  
Db 421 TCCAGTCCCGATCCGTTATTTAACTCTTAAAACTGCGTTTGCCACTGATTAACGCG 480  
Qy 481 AACGTGACTGACGCGATCTCTCAGCAGGGAGAGGGTCAATTGCTGACTTTACCGGGCAT 540  
Db 481 AACGTGACTGACGCGATCTCTCAGCAGGGAGAGGGTCAATTGCTGACTTTACCGGGCAT 540  
Qy 541 CCGCAAAACGGCGTTTGCGCAACTGGAACGGGTGCTTAATTTTCGCAATCAACTGTGCG 600  
Db 541 TATCAAAACGGCGTTTGCGCAACTGGAACGGGTGCTTAATTTTCGCAATCAACTGTGCG 600  
Qy 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTTCAATTAACGAGGCAATTACCATCGAACTC 660  
Db 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTTCAATTAACGAGGCAATTACCATCGAACTC 660  
Qy 661 AAGGTAGCGCGCGCAATGTTCTTATTAACGGTGGGTAAAGCTTGGCATCAATGCTGACG 720  
Db 661 AAGGTAGCGCGCGCAATGTTCTTATTAACGGTGGGTAAAGCTTGGCATCAATGCTGACG 720  
Qy 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCCGAGCCGGGATGGGGAAGATCACG 780  
Db 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCCGAGCCGGGATGGGGAAGATCACG 780  
Qy 781 GATTACACCAAGTGAACACCTTGCTAAAGTTTGCAATACGCGCAATTTAATTTGCTACAA 840  
Db 781 GATTACACCAAGTGAACACCTTGCTAAAGTTTGCAATACGCGCAATTTAATTTGCTACAA 840  
Qy 841 CGCAGCCGAGAGGTGGCCGCGACGCCGCCACCCCGTTATTAAATTGATCAAGACAGCG 900  
Db 841 CGCAGCCGAGAGGTGGCCGCGACGCCGCCACCCCGTTATTAAATTGATCAAGACAGCG 900  
Qy 901 TTTAGCCGCGATTCACCGGAAAAACGGCGGTATGTTGTCATTAACCATTCAGTGTG 960  
Db 901 TTTAGCCGCGATTCACCGGAAAAACGGCGGTATGTTGTCATTAACCATTCAGTGTG 960  
Qy 961 TTTATGCGCGGACAGATCTAATCTGCAAAATCTCGGCGGCGCATGAGCTCACTGCG 1020  
Db 961 TTTATGCGCGGACAGATCTAATCTGCAAAATCTCGGCGGCGCATGAGCTCACTGCG 1020  
Qy 1021 ACCGTTCCCGGTACGCCGATTAACAGCCGCCAGGTGTGAATGTTTGAACGCTGG 1080  
Db 1021 ACCGTTCCCGGTACGCCGATTAACAGCCGCCAGGTGTGAATGTTTGAACGCTGG 1080  
Qy 1081 CGTCCGCTAAGCATTAACGCCAGTGAATCAGGTTTCCGCTTCCAGACTTTCAG 1140  
Db 1081 CGTCCGCTAAGCATTAACGCCAGTGAATCAGGTTTCCGCTTCCAGACTTTCAG 1140  
Qy 1141 CAGATGCGTATTAACGCCGCTGTCAATTAATACGCCGCCGAGAGGTGAACGTGACC 1200  
Db 1141 CAGATGCGTATTAACGCCGCTGTCAATTAATACGCCGCCGAGAGGTGAACGTGACC 1200  
Qy 1201 CTGGCAGATGTGAAGAGCAATGCGCAGGGCATGTGTTGCGCAGGTTTACGCAA 1260  
Db 1201 CTGGCAGATGTGAAGAGCAATGCGCAGGGCATGTGTTGCGCAGGTTTACGCAA 1260  
Qy 1261 ATGCTGAATGAAGCAGCATACCGGCTTGAGGTTTG 1296  
Db 1261 ATGCTGAATGAAGCAGCATACCGGCTTGAGGTTTG 1296

Search completed: February 10, 2005, 18:17:13  
Job time : 956.54 secs



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33; Search time 5631.01 Seconds  
(without alignments)  
8780.927 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_12653\_13951

Perfect score: 1299

Sequence: 1 atgaagaccatcttaacc.....taccgcctgcagttgttaa 1299

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: EST.\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hic.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gest1.\*

9: gb\_gest2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	614.4	47.3	616	BU074127	BU074127 BU074127
2	580.8	44.7	853	CL662734	CL662734 PRI0142b
3	536.2	41.3	746	CA093060	CA093060 SCCCL200
4	529	40.7	529	BU713770	BU713770 SUABUC02
5	488.8	37.6	595	CP325092	CP325092 JMT1--05-
6	483.8	37.2	895	CN762997	CN762997 ID0AA5DC
7	445.8	34.3	864	CN757565	CN757565 ID0AA1DF
8	441.8	34.0	864	CN757565	CN757565 ID0AA1DF
9	386.2	29.7	393	BE520240	BE520240 M1B128TM
10	383	29.5	383	AM036132	AM036132 EST274510
11	383	29.5	383	AM036134	AM036134 EST274510
12	353.4	27.2	531	BU713769	BU713769 SUABUC01
13	304.6	23.4	354	AA545747	AA545747 HBMSF1B4-
14	294.8	22.7	706	CN759004	CN759004 ID0AA24B
15	256	19.7	707	BU619443	BU619443 ID0AA24B
16	216.8	16.7	320	BE436403	BE436403 EST407481
17	208	16.0	208	BM409016	BM409016 EST583343
18	208	16.0	208	BM412806	BM412806 EST583733
19	198.8	15.3	546	BU040796	BU040796 BU040796
20	169.4	13.0	179	BE461872	BE461872 EST13291
21	162.4	12.5	657	BG457132	BG457132 NF100C05P
22	160.8	12.4	508	BU030203	BU030203 BU030203
23	141.6	10.9	254	BE520241	BE520241 M1B128TM
24	103.4	8.0	769	CO386885	CO386885 AGENCOURT

C 25	54.6	4.2	821	9	CL665831	CL665831 PRI014d_E
C 26	54.6	4.2	842	7	CL665745	CL665745 PRI014c_E
C 27	43.8	3.4	392	7	H60613	H60613 Yr4d12..81
C 28	42.2	3.2	733	4	BI152280	BI152280 602917706
C 29	40.8	3.1	346	5	BY053637	BY053637 BY053637
C 30	40.8	3.1	350	5	BY041909	BY041909 BY041909
C 31	40.8	3.1	365	5	BY056036	BY056036 BY056036
C 32	40.8	3.1	389	6	CA566872	CA566872 K0406H12-
C 33	40.8	3.1	431	1	AI553154	AI553154 v107b12..Y
C 34	40.8	3.1	539	5	BX517075	BX517075 BX517075
C 35	40.8	3.1	602	2	BB610221	BB610221 BB610221
C 36	40.8	3.1	607	1	AA521521	AA521521 v107b12..T
C 37	40.8	3.1	632	6	BY738708	BY738708 BY738708
C 38	40.8	3.1	646	6	BY737657	BY737657 BY737657
C 39	40.8	3.1	2931	3	AK075739	AK075739 Mus muscu
C 40	39.2	3.0	379	2	BE289814	BE289814 601089743
C 41	39.2	3.0	411	6	CA880619	CA880619 K0984D03-
C 42	39.2	3.0	424	5	BY261762	BY261762 BY261762
C 43	39.2	3.0	434	6	BY238933	BY238933 BY238933
C 44	39.2	3.0	436	6	CA895377	CA895377 B0191P11-
C 45	39.2	3.0	485	6	CA888817	CA888817 B0148G10-

## ALIGNMENTS

RESULT 1  
BU074127/c 616 bp mRNA linear EST 29-SEP-2003  
BU074127 NIBB Mochii normalized Xenopus tailbud library Xenopus  
DEFINITION  
laevis cDNA clone XL090j06 5', mRNA sequence.  
BU074127  
BU074127.1 GI:17504316

KEYWORDS  
EST.  
Xenopus laevis (African clawed frog)

ORGANISM  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
Xenopodinae; Xenopus; Xenopus.

REFERENCE  
1 (bases 1 to 616)

AUTHORS  
Kikayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and  
Kohara,Y.

TITLE  
Expressed genes in X. laevis embryo

QY 324 TCGTATGTCGACGAGCGTACCCGTTAAACAGCGCAAGCTTTCGCCGCCGCGGCTGGCAC 383  
DB 556 TCGTATGTCGACGAGCGTACCCGTTAAACAGCGCAAGCTTTCGCCGCCGCGGCTGGCAC 497  
QY 384 TGAAGTGTGAATTAACCGTACATACCGAGGAGATACGTCAGTCCCGATCCGTTATTTAA 443  
DB 496 TGAAGTGTGAATTAACCGTACATACCGAGGAGATACGTCAGTCCCGATCCGTTATTTAA 437  
QY 444 TCCCTTAAACAACTGCGTTTGGCACTGATTAACCGGAACGTCAGTGAACGATCCTCG 503  
DB 436 TCCCTTAAACAACTGCGTTTGGCACTGATTAACCGGAACGTCAGTGAACGATCCTCG 377  
QY 504 CAGGCGAGAGGAGTCAATTCGTCATTTACCGGCGCATCGGCAACAGGCGCTTTCGGAAT 563  
DB 376 CAGGCGAGAGGAGTCAATTCGTCATTTACCGGCGCATCGGCAACAGGCGCTTTCGGAAT 317  
QY 564 GGAACGCGGCTTAATTTCCGCAATCAAACTTGTGCTTAAACGTGAACAGAGCA 623  
DB 316 GGAACGCGGCTTAATTTCCGCAATCAAACTTGTGCTTAAACGTGAACAGAGCA 257  
QY 624 AAGCTGTTCATTAACGACGAGCAATTAACATCGAATCAAGTGAGCGCGCAATGTCTC 683  
DB 256 AAGCTGTTCATTAACGACGAGCAATTAACATCGAATCAAGTGAGCGCGCAATGTCTC 197  
QY 684 ATTTAACCGGCTGCGTAAAGCTCGCATCAATGTCAGAGGAGATTTTCTCTGCAACAGC 743  
DB 196 ATTTAACCGGCTGCGTAAAGCTCGCATCAATGTCAGAGGAGATTTTCTCTGCAACAGC 137  
QY 744 ACAGGAGATGCGGAGCGGAGTGGGAGAGATCAACGATCAACACGATGAGAACCTT 803  
DB 136 ACTGGAGATGCGGAGCGGAGTGGGAGAGATCAACGATCAACACGATGAGAACCTT 77  
QY 804 GCTAAGTTTGCATTAACGCGCAATTTTATTGCTCAACAGCAGCGCAGAGTTGCCGAG 863  
DB 76 GCTAAGTTTGCATTAACGCGCAATTTTATTGCTCAACAGCAGCGCAGAGTTGCCGAG 17  
QY 864 CCGCGCCACCCCGTTA 879  
DB 16 CCGCGCCACCCCGTTA 1

RESULT 2  
LOCUS CL662734 853 bp DNA linear GSS 09-JUL-2004  
DEFINITION PRI0142b.B1 - PRI0142b.B21 (853) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.  
ACCESSION CL662734  
VERSION CL662734.1 GI:50150877  
KEYWORDS GSS.  
SOURCE Pristionchus pacificus  
ORGANISM Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasteridae; Neodiplogasteridae; Pristionchus.  
REFERENCE 1 (bases 1 to 853)  
AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
TITLE AppaB: an Acedb database for the nematode satellite organism Pristionchus pacificus  
JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)  
COMMENT Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.  
Seq primer: 17  
Classes: fosmid ends.  
FEATURES  
Source 1..853 Location/Qualifiers

/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_id="Mixed stage fosmid library of P. pacificus var. California"  
/note="Vector: pBfi05-5 Fosmid vector"

ORIGIN

Query Match 44.7%; Score 580.8; DB 9; Length 853;  
Best Local Similarity 98.8%; Pred. No. 5.4e-171;  
Matches 585; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 708 ATCAATGCTGACGGAGATATTCTCCGCAACAGACAGAGGAATGCGAGCGGAGT 767  
DB 1 ATCAATGCTGACGGAGATATTCTCCGCAACAGACAGAGGAATGCGAGCGGAGT 60  
QY 768 GGGAGGATCAACCGATTCAACAGTGGAAACCTTGTGATTAACGCGCAATT 827  
DB 61 GGGAGGATCAACCGATTCAACAGTGGAAACCTTGTGATTAACGCGCAATT 120  
QY 828 TTATTGCTACAGGACGCGAGAGTGGCCCGGCGCCACCCGTTATTAGATT 887  
DB 121 TTATTGCTACAGGACGCGAGAGTGGCCCGGCGCCACCCGTTATTAGATT 180  
QY 888 GATCAAGACAGGTTGACGCGCCCATCCACGCAAAAACAGGCGATGATGACATTAC 947  
DB 181 GATCAAGACAGGTTGACGCGCCCATCCACGCAAAAACAGGCGATGATGACATTAC 240  
QY 948 CACTTCAAGTGTGTTTATGCGCGGACAGATACTATCTGCAAACTTCGCGGCGCACT 1007  
DB 241 CACTTCAAGTGTGTTTATGCGCGGACAGATACTATCTGCAAACTTCGCGGCGCACT 300  
QY 1008 GAGCTCAACTGAGAGCTTCCCGGTACAGCGGATTAACAGCGGATGGAATGCT 1067  
DB 301 GAGCTCAACTGAGAGCTTCCCGGTACAGCGGATTAACAGCGGATGGAATGCT 360  
QY 1068 GTTTGAACGTCGCGCGTGAAGGATTAAGCAGATGATTCAGTTTGTGCTGCTT 1127  
DB 361 GTTTGAACGTCGCGCGTGAAGGATTAAGCAGATGATTCAGTTTGTGCTGCTT 420  
QY 1128 CCAGACTTTTACAGCAGATGCGTGAATAAAGCGCGTGCATTAATAACCGCGCGGAGA 1187  
DB 421 CCAGACTTTTACAGCAGATGCGTGAATAAAGCGCGTGCATTAATAACCGCGCGGAGA 480  
QY 1188 GGTGAACCTGACCTCGGAGAGATGGAAGAGGAATGCGAGGCGATGTTGCTGTC 1247  
DB 481 GGTGAACCTGACCTCGGAGAGATGGAAGAGGAATGCGAGGCGATGTTGCTGTC 540  
QY 1248 AGTTTTCGCAAAATCGTAATGAAGACAGCATACCGCTTGCAATTGTGA 1299  
DB 541 AGTTTTCGCAAAATCGTAATGAAGACAGCATACCGCGTGCATTTGTGA 592

RESULT 3  
LOCUS CA093060 746 bp mRNA linear EST 23-SEP-2003  
DEFINITION SCCCL2001D10.b CL2 Saccharum officinarum cDNA clone SCCCL2001D10 3', mRNA sequence.  
ACCESSION CA093060  
VERSION CA093060.1 GI:34946367  
KEYWORDS EST.  
SOURCE Saccharum officinarum  
ORGANISM Saccharum officinarum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum complex.  
REFERENCE 1 (bases 1 to 746)  
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
TITLE The libraries that made SUCEST  
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
COMMENT Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica  
Universidade Estadual de Campinas  
Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
Tel: 55 19 3768 1137  
Fax: 55 19 3768 1089  
Email: pattud@unicamp.br  
Clone distribution: clone distribution information can be found  
through the Brazilian Clone Collection Center (BCCC) at  
http://www.bcccenter.fcav.unesp.br  
Plate: 001 row: D column: 10  
Seq primer: M13/Forward primer.  
Location/Qualifiers

## FEATURES

source

1. /746  
/organism="Saccharum officinarum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCCCL2001D10"  
/lab\_host="X11Blue MRP"  
/clone\_lib="CL2"  
/note="Organ: Pool of sugarcane calli submitted to low  
(40C) and high (37 C) temperature stress; Vector:  
pBluscript; Site 1: EcoRI; Site 2: XhoI; An  
unidirectional cDNA library generated from [Pool of  
sugarcane calli submitted to low (40C) and high (37 C)  
temperature stress]. cDNA was prepared from polyA+ mRNA  
using ZAP - cDNA Synthesis Kit (Stratagene). The  
double-strand cDNAs were fractionated in a sepharose CL-2B  
40cm-columns and fragments siting between 0.8 and 1.5 Kb  
were directionally cloned into the vector. Details of  
each source of RNA and library construction can be  
obtained at http://bucest.lad.fc.unicamp.br/public"

## ORIGIN

Query Match 41.3%; Score 536.2; DB 6; Length 746;  
Best Local Similarity 97.9%; Pred. NO. 5.8e-157;  
Matches 552; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

595 TTGTGCTTAAAGCTGAGAAACAGAGCAAGAGCTGTTTCATTAAAGCAGGATTACATCG 654  
184 TTGCGTACTGTCAGTGAAGAAAGAGCAAGAGCTGTTTCATTAAAGCAGGATTACATCG 243  
655 GAACTCAAGGTGAGCGCGCAATGTTTCATTAAAGCAGGATTACATCG 714  
244 GAACTCAAGGTGAGCGCGCAATGTTTCATTAAAGCAGGATTACATCG 303  
715 CTGACGAGATATTTCTCCGCAACAGCAAGGAAATCCGCGCGGGTGGGAGG 774  
304 CTGACGAGATATTTCTCCGCAACAGCAAGGAAATCCGCGCGGGTGGGAGG 363  
775 ATCAGCATTCACACAGTGAACACCTTGTCTAAGTTTGCATTAACGCGCATATTTATTTG 834  
364 ATCAGCATTCACACAGTGAACACCTTGTCTAAGTTTGCATTAACGCGCATATTTATTTG 423  
835 CTACACGACGACGACAGAGTTTCCCGCAGCGCGCACCCCGTTATTAGATTGATCAG 894  
424 CTACACGACGACGACAGAGTTTCCCGCAGCGCGCACCCCGTTATTAGATTGATCAG 483  
895 ACAGCGTTGAGCGCGCGCATCCACCCCAAAAAGAGCGTATGATGATTCACCATTTCA 954  
484 ACAGCGTTGAGCGCGCGCATCCACCCCAAAAAGAGCGTATGATGATTCACCATTTCA 543  
955 GTGCTGTTTATCGCGCGACGATATCTAATCTGCAATCTCGGCGCGGCACTGGAGCTC 1014  
544 GTGCTGTTTATCGCGCGACGATATCTAATCTGCAATCTCGGCGCGGCACTGGAGCTC 603  
1015 AACTGAGAGCTTCCCGGTGACCGCGATTAACGCGCGCAAGGTGTAATCTGTTTGA 1074  
604 AACTGAGAGCTTCCCGGTGACCGCGATTAACGCGCGCAAGGTGTAATCTGTTTGA 662  
1075 CGCTGCGCTCGCTTAAGGATTAACGCGAGTGATTCAGGTTTGGCTGCTTCAGACT 1134  
663 CGCTGCGCTCGCTTAAGGATTAACGCGAGTGATTCAGGTTTGGCTGCTTCAGACT 722

QY 1135 TTACAGCAGATCGTGATTAACG 1158  
Db 723 TTACAGCANATCGTGATTAACG 746

## RESULT 4

BU713770

## LOCUS

DEFINITION

BU713770 529 bp mRNA linear EST 23-OCT-2003  
SUABUG02 Adult SJC 7/94 Schistosoma japonicum cDNA similar to  
pddlIDK1A Chain A, Crystal Structure Of Phytate Complex Of  
Escherichia Coli Phytase At Ph 6.6. Phytate Is Bound With Its  
3-Phosphate In The Active Site. H92+ Cation Acts As An  
Intermolecular Bridge, mRNA sequence.

## ACCESSION

VERSION

## SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

COMMENT

CONTACT

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn.

Location/Qualifiers

1. 529

/organism="Schistosoma japonicum"

/mol\_type="mRNA"

/strain="Chinese (Anhui) strain"

/db\_xref="taxon:6182"

/sex="Male and female"

/tissue\_type="Whole body"

/dev stage="Adult worms"

/lab host="Mouse and rabbit"

/clone\_lib="Adult SJC 7/94"

/note="Vector: Lambda ZAP-II XR; Site 1: EcoRI; Site 2:

XhoI I; Several hundred adult Schistosoma japonicum

from the mesenteries of experimentally infected mice and

rabbits at the Queensland Institute of Medical Research,

Brisbane, Australia (QIMR), and stored for several months

in liquid nitrogen. Subsequently, mRNA was isolated at the

QIMR from lypates of these worms by oligo dt

chromatography, using a kit from Pharmacia. The mRNA was

then shipped to Clontech, Palo Alto, CA, USA, who

constructed a cDNA library. First strand synthesis was

primed with an oligo-dt-XhoI primer and synthesized using

M-MuV reverse transcriptase. Second strand synthesis was

accomplished with RNase H and T4 DNA polymerase. The

double stranded cDNA was ligated to EcoRI linkers,

digested with EcoRI and XhoI, and ligated into the

phagemid vector lambda ZAP II XR. After construction of

this directional library by Clontech, it was returned to

the QIMR. During analysis of the library at the QIMR, we

have found that a small percentage, 2% to 3%, of the

clones contain inserts that appear to be highly homologous

to sequences from salmonid fishes, as determined by

homology comparisons using BLAST and by Southern

hybridization analysis to genomic DNA from salmon (Sigma

Chemical Co., St. Louis, MO) under stringent washing

conditions. The remainder of the clones appear to contain  
S. japonicum sequences."

Query Match 40.7%; Score 529; DB 5; Length 529;  
Best Local Similarity 100.0%; Pred. No. 9.8e-155;  
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 GCCGACAACTCTCATTAACCGGTGGGTAAAGCTTCGATCAATGCTGACGAGATATTT 729

DB 1 GCCGAAAGTGTCTCATTAACCGGTGGGTAAAGCTTCGATCAATGCTGACGAGATATTT 60

QY 730 CTCCTGCAACAAGCAGAGGAATGCGGAGCCGGGTGGGGAAGATCAACGATTCACAC 789

DB 61 CTCCTGCAACAAGCAGAGGAATGCGGAGCCGGGTGGGGAAGATCAACGATTCACAC 120

QY 790 CAGTGGAACCTTGTCTTAAGTTTGCAATACGGCAATTTTATTGCTACAGCAGCCCA 849

DB 121 CAGTGGAACCTTGTCTTAAGTTTGCAATACGGCAATTTTATTGCTACAGCAGCCCA 180

QY 850 GAGGTGCGCGGAGCGCGCCACCGGTTATTAAGATTAATCAAGACAGGTTGAGCGCC 909

DB 181 GAGGTGCGCGGAGCGCGCCACCGGTTATTAAGATTAATCAAGACAGGTTGAGCGCC 240

QY 910 CATCCACCGCAAAAAGAGCGATGATGACATTAACCATTCAGTGTGTTATCGCC 969

DB 241 CATCCACCGCAAAAAGAGCGATGATGACATTAACCATTCAGTGTGTTATCGCC 300

QY 970 GGAACGATACTAATCTGGCAATCTCGGCGGCGCACTGAGCTCAACTGAGCGTTCC 1029

DB 301 GGAACGATACTAATCTGGCAATCTCGGCGGCGCACTGAGCTCAACTGAGCGTTCC 360

QY 1030 GGTGACCGCGGATTAACACGCGCGAGGTGAACTGTGTGTTGAAGCGGTGCGGTA 1089

DB 361 GGTGACCGCGGATTAACACGCGCGAGGTGAACTGTGTGTTGAAGCGGTGCGGTA 420

QY 1090 AAGGATTAACGACGAGGATTCAGTGTGCTGTGCTTCAGACTTAAGAGATGCGT 1149

DB 421 AAGGATTAACGACGAGGATTCAGTGTGCTGTGCTTCAGACTTAAGAGATGCGT 480

QY 1150 GATTAACGCGCGTGTCTTAATTAATACGCGCGCGGAGAGGTGAACCTGA 1198

DB 481 GATTAACGCGCGTGTCTTAATTAATACGCGCGCGGAGAGGTGAACCTGA 529

RESULT 5  
CF326092/c 595 bp mRNA linear EST 18-AUG-2003  
LOCUS JMT1--05-B09-g1 ACUMT-overexpressing transgenic rice lambda phage  
DEFINITION cDNA library (JMT1) Oryza sativa (japonica cultivar-group) cDNA  
clone JMT1--05-B09, mRNA sequence.

ACCESSION CF326092  
VERSION CF326092.1 GI:33800445  
KEYWORDS EST.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 595)  
AUTHORS Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,  
Song,S.I., Kim,J.K., Kim,Y.-K. and Nahm,B.H.  
TITLE Large-scale Sequencing Analysis of Rice ESTs  
JOURNAL Unpublished (2003)  
COMMENT Contact: Nahm B.H.  
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division  
of Bioscience and Bioinformatics, Myongji University  
Yongin, Kyeonggi, Korea  
Tel: 82 31 330 6193  
Fax: 82 31 321 6355  
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES  
Location/Qualifiers  
1..595  
source

/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cufix="Nackdong"  
/db\_xref="taxon:39947"  
/clone="JMT1--05-B09"  
/tissue\_type="leaf"  
/dev\_stage="14 days after germination"  
/lab\_host="E.coli SDR"  
/clone\_lib="ACUMT-overexpressing transgenic rice lambda  
phage cDNA library (JMT1)"  
/note="Vector: pBluescript SK(+), Site 1: EcoRI, Site 2:  
XhoI; cDNA was inserted into lambda Uni-Zap XR vector at 5'  
end with EcoRI and 3' end with XhoI site. mRNA was  
prepared from Arabidopsis Jasmonate Carboxyl  
methyltransferase overexpression line."

Query Match 37.6%; Score 488.8; DB 7; Length 595;  
Best Local Similarity 99.6%; Pred. No. 4.6e-142;  
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 808 AGTTGCAATACGGCCATTTTATTGCTACACAGCAGAGGTGCGCGAGCGCC 867

DB 595 AGTTGCAATACGGCCATTTTATTGCTACACAGCAGAGGTGCGCGAGCGCC 536

QY 868 GCCACCCGGTTATTAATTTGATCAAGACAGGTTGAGCGCCATCCACCGCAAAAAG 927

DB 535 GCCACCCGGTTATTAATTTGATCAAGACAGGTTGAGCGCCATCCACCGCAAAAAG 476

QY 928 GCGATGATGATCAATACCACTTCACTGCTGTTATTCGCGGACAGATATCTATCTG 987

DB 475 GCGATGATGATCAATACCACTTCACTGCTGTTATTCGCGGACAGATATCTATCTG 416

QY 988 GGAATCTTCGGGGGCGACTGAGCTCACTGAGAGCTTCGCCGTCACCGGATTAACG 1047

DB 415 GGAATCTTCGGGGGCGACTGAGCTCACTGAGAGCTTCGCCGTCACCGGATTAACG 356

QY 1048 CCGCAGGTGATGAACTGATGTTGAACGCTGCGCTGCGATTAAGATTAACGACGATG 1107

DB 355 CCGCAGGTGATGAACTGATGTTGAACGCTGCGCTGCGATTAAGATTAACGACGATG 296

QY 1108 ATTCAAGTTTGGCTGCTGCTTCCAGACTTACAGCAGATGCGTATTAACGCGCTGCA 1167

DB 295 ATTCAAGTTTGGCTGCTGCTTCCAGACTTACAGCAGATGCGTATTAACGCGCTGCA 236

QY 1168 TTAATATGCGCGCCCGGAGAGGTGAACCTTCGAGAGATGGAAGGGAATCG 1227

DB 235 TTAATATGCGCGCCCGGAGAGGTGAACCTTCGAGAGATGGAAGGGAATCG 176

QY 1228 CAGGCGATGTTGCTGTTGAGAGTTTACGAAATCGTAATGAAGCAGCATACCGCT 1287

DB 175 CAGGCGATGTTGCTGTTGAGAGTTTACGAAATCGTAATGAAGCAGCATACCGCT 116

QY 1288 TGCAGTTGTAA 1299  
DB 115 TGCAGTTGTAA 104

RESULT 6  
CN762997/c 895 bp mRNA linear EST 20-MAY-2004  
LOCUS ID0AA5DC09BRI AgMS Acyrthosiphon pisum cDNA clone ID0AA5DC09 5',  
DEFINITION mRNA sequence.  
ACCESSION CN762997  
VERSION CN762997.1 GI:47536920  
KEYWORDS EST.  
SOURCE Acyrthosiphon pisum (pea aphid)  
ORGANISM Acyrthosiphon pisum  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Psyllota;  
Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;  
Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

REFERENCE 1 (bases 1 to 895)  
AUTHORS Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,

TITLE Stern, D., Tagu, D. and Wincker, P.  
An expressed sequence tags database for the pea aphid *Acyrthosiphon pisum*  
JOURNAL Unpublished (2004)  
COMMENT Contact: D. Tagu  
INRA Rennes  
UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France  
Tel: +33.2.23.48.51.65  
Fax: +33.2.23.48.51.50  
Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers  
FORWARD: CAGGAAACAGCTATGACC  
Plate: 5  
Location/Qualifiers  
1. .895  
/organism="Acyrthosiphon pisum"  
/mol\_type="mRNA"  
/cultivar="developmentstage"  
/db\_xref="taxon:7029"  
/clone="ID0AAA5DC09"  
/tissue\_type="whole insect"  
/dev\_stage="nymphs and adults (parthenogenetic females)"  
/lab\_host="XLI-Blue"  
/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University; Soil conditions: Soil; Sowing date: 01/06/1999; Harvesting date: 01/06/1999; Stress date: no stress; Description: Aphids inoculated on one-week old *Vicia faba* under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 °C)"

## ORIGIN

Query Match 37.2%; Score 483.8; DB 7; Length 895;  
Best Local Similarity 99.6%; Pred. No. 1.9e-140;  
Matches 485; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

813 GCATACGCGCAATTTATTTGCTACACGACCGCCAGAGTTGCCGCGACCGCCGAC 872  
895 GCATACGCGCAATTTATTTGCTACACGACCGCCAGAGTTGCCGCGACCGCCGAC 836  
873 CCCGTTTATGATTTGATCAAGACAGCGTTGACCGCCCATGACCGCAAAACAGGCGTA 932  
835 CCCGTTTATGATTTGATCAAGACAGCGTTGACCGCCCATGACCGCAAAACAGGCGTA 776  
933 TGATGATACATTACCACTTCAGTGTCTTTATGCGCGGACAGATATTAATCTGGCAAA 992  
775 TGATGATACATTACCACTTCAGTGTCTTTATGCGCGGACAGATATTAATCTGGCAAA 716  
993 TCTGCGGCGGCGACTGAGCTCAACTGAGCGCTTCCGCGTACGCGGATTAACAGCGCGCC 1052  
715 TCTGCGGCGGCGACTGAGCTCAACTGAGCGCTTCCGCGTACGCGGATTAACAGCGCGCC 656  
1053 AGGTGATGAATGATGTTTGAACGCTGCGCTGCGTTAAGGATTAACAGCGCGGATTTCA 1112  
655 AGGTGATGAATGATGTTTGAACGCTGCGCTGCGTTAAGGATTAACAGCGCGGATTTCA 596  
1113 GGTTCGCTGCTCTTCAGACTTTTACAGCAGATGCGTATTAACAGCGCGCTGTCTTAA 1172  
595 GGTTCGCTGCTCTTCAGACTTTTACAGCAGATGCGTATTAACAGCGCGCTGTCTTAA 536  
1173 TACCGCGCGCGAGAGTGAATCTGACCTTGGCAGAGATGTGAAGAGCAATTCGCGAGG 1232  
535 TACCGCGCGCGAGAGTGAATCTGACCTTGGCAGAGATGTGAAGAGCAATTCGCGAGG 476  
1233 CATGTGCTGTTGGAGGTTTAAAGCAATGCTGAATGAAGACGATACCGCGCTTGAG 1292

Db 475 CATGTGCTGTTGGAGGTTTAAAGCAATGCTGAATGAAGACGATACCGCGCTGAG 416  
Qy 1293 TTTGTAA 1299  
Db 415 TTTGTAA 409

RESULT 7  
CN754382/c 868 bp mRNA linear EST 19-MAY-2004  
LOCUS ID0AAA12DE01RM1 APMS Acyrthosiphon pisum cdna clone ID0AAA12DE01  
DEFINITION 5', mRNA sequence.  
ACCESSION CN754382  
VERSION CN754382.1 GI:47519379  
KEYWORDS EST.  
SOURCE Acyrthosiphon pisum (pea aphid)  
ORGANISM Acyrthosiphon pisum  
Acyrthosiphon pisum  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

REFERENCE 1 (bases 1 to 868)  
Hunter, W., Martinez-Torres, D., Raabe, Y., Sabater-Munoz, B., Stern, D., Tagu, D. and Wincker, P.  
An expressed sequence tags database for the pea aphid *Acyrthosiphon pisum*  
JOURNAL Unpublished (2004)  
COMMENT Contact: D. Tagu  
INRA Rennes  
UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France  
Tel: +33.2.23.48.51.65  
Fax: +33.2.23.48.51.50  
Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

## FEATURES

source  
Location/Qualifiers  
1. .868  
/organism="Acyrthosiphon pisum"  
/mol\_type="mRNA"  
/cultivar="developmentstage"  
/db\_xref="taxon:7029"  
/clone="ID0AAA12DE01"  
/tissue\_type="whole insect"  
/dev\_stage="nymphs and adults (parthenogenetic females)"  
/lab\_host="XLI-Blue"  
/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University; Soil conditions: Soil; Sowing date: 01/06/1999; Harvesting date: 01/06/1999; Stress date: no stress; Description: Aphids inoculated on one-week old *Vicia faba* under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 °C)"

## ORIGIN

Query Match 34.3%; Score 445.8; DB 7; Length 868;  
Best Local Similarity 99.3%; Pred. No. 1.8e-128;  
Matches 458; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

839 AACGACGCGCAGAGTTGCGCGGCGCCACCGCTTATTTAGATTGATCAAGACAG 898  
868 AACGACGCGCAGAGTTGCGCGGCGCCACCGCTTATTTAGATTGATCAAGACAG 810  
899 CGTTAGCGCCCATTCACCGCAAAACAGGCGTATGTTGACATTAACCATTCAGTGC 958  
809 CGTTAGCGCCCATTCACCGCAAAACAGGCGTATGTTGACATTAACCATTCAGTGC 750

QY 959 TGTATTGCGCGGACACGATCTAATCTGGCAATCTCGCGCGCGACTGAGCTCAACT 1018  
 749 TGTATTGCGCGGACACGATCTAATCTGGCAATCTCGCGCGCGACTGAGCTCAACT 690  
 QY 1019 GGAAGCTTCCCGGTGACCGCGATTAACGCCCGCGAGTGTGTAAGTGTGTAAGCT 1078  
 689 GGAAGCTTCCCGGTGACCGCGATTAACGCCCGCGAGTGTGTAAGTGTGTAAGCT 630  
 Db 1079 GGGGTGGGTAAAGCGATTAACGCGAGTGTGTAAGTGTGTAAGTGTGTAAGCT 1138  
 629 GGGGTGGGTAAAGCGATTAACGCGAGTGTGTAAGTGTGTAAGTGTGTAAGCT 570  
 QY 1139 AGCAGATGCGTGAATAAAGCGCGGTGATTAATACCGCGCGAGTGTGTAAGTGTGTAAGCT 1198  
 569 AGCAGATGCGTGAATAAAGCGCGGTGATTAATACCGCGCGAGTGTGTAAGTGTGTAAGCT 510  
 Db 1199 CCCTGGCAGATGTGAAGAGCGAAATGCGCGAGTGTGTAAGTGTGTAAGTGTGTAAGCT 1258  
 509 CCCTGGCAGATGTGAAGAGCGAAATGCGCGAGTGTGTAAGTGTGTAAGTGTGTAAGCT 450  
 QY 1259 AAATCGTGAATGAAGACGCGATACCGCGCTTGCAATTGTGA 1299  
 449 AAATCGTGAATGAAGACGCGATACCGCGCTTGCAATTGTGA 409  
 Db

RESULT 8  
 CN757565/c 864 bp mRNA linear EST 20-MAY-2004  
 LOCUS ID0AA1D09RM1 ApMs Acyrthosiphon pisum cDNA clone ID0AA1D09 5',  
 DEFINITION mRNA sequence.  
 ACCESSION CN757565  
 VERSION CN757565.1 GI:47531488  
 KEYWORDS EST.  
 SOURCE Acyrthosiphon pisum (pea aphid)  
 ORGANISM Acyrthosiphon pisum  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;  
 Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.  
 1 (bases 1 to 864)  
 Hunter, W., Martinez-Torres, D., Rahbe, Y., Sabater-Munoz, B.,  
 Stern, D., Tagu, D. and Winkler, P.  
 An expressed sequence tags database for the pea aphid Acyrthosiphon

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 UNPUBLISHED (2004)  
 Contact: D. Tagu  
 INRA Rennes  
 UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France  
 Tel: +33.2.23.48.51.65  
 Fax: +33.2.23.48.51.50

FEATURES  
 source  
 1..864  
 /organism="Acyrthosiphon pisum"  
 /mol\_type="mRNA"  
 /cultivar="developmentstage"  
 /db\_xref="taxon:7029"  
 /clone="ID0AA1D09"  
 /issue\_type="whole insect"  
 /dev\_stage="nymphs and adults (parthenogenetic females)"  
 /lab\_host="Xyl-Blue"  
 /clone\_lib="ApMs"  
 /note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI;  
 Sample name: ID0AAA; Plant growth place: Department of  
 Ecology & Evolutionary Biology, Princeton University;  
 Soil conditions: Soil; Sowing date: 01/06/1999;  
 Harvesting date: 01/06/1999; Stress date: no stress;  
 Description: Aphids inoculated on one-week old *Vicia faba*

## ORIGIN

Query Match 34.0%; Score 441.8; DB 7; Length 864;  
 Query Similarity 99.3%; Pred. No. 3,3e-127;  
 Matches 454; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

under non-sterile conditions. All parthenogenetic stages  
 and both winged and wingless adults were collected for  
 library construction. ; experimental condition: long  
 photoperiod (16-hr light/8-hr dark at 18 °C)

QY 843 CAGCCAGAGTGTCCCGCGAGCCCGCGACCCCGTATTAGATTGATCAAGACAGCTT 902  
 864 CAGCCAGAGTGTCCCGCGAGCCCGCGACCCCGTATTAGATTGATCAAGACAGCTT 806  
 Db 903 GAGCCCGCATCAACCGCAAAAGAGCGTATGATGATCAATTACCACTTCACTGCTGTT 746  
 805 GAGCCCGCATCAACCGCAAAAGAGCGTATGATGATCAATTACCACTTCACTGCTGTT 746  
 QY 963 TATCGCCGACACGATCTAATCTGCGCAATCTCGCGCGCGCACTGAGCTCAACTGGAC 1022  
 745 TATCGCCGACACGATCTAATCTGCGCAATCTCGCGCGCGCACTGAGCTCAACTGGAC 686  
 Db 1023 GCTTCCCGGTAGCGCGGATTAACAGCCCGCAGGTGTGAATGTTTGAACGCTGGCG 1082  
 685 GCTTCCCGGTAGCGCGGATTAACAGCCCGCAGGTGTGAATGTTTGAACGCTGGCG 626  
 QY 1083 TCGGCTAAGCATTAACAGCCAGTGAATCAAGTTCCTGCTCTTCCAGCTTTACGCA 1142  
 625 TCGGCTAAGCATTAACAGCCAGTGAATCAAGTTCCTGCTCTTCCAGCTTTACGCA 566  
 Db 1143 GATGCGTATTAAGCGCGCTGTGATTAATAGCGCGCGCGAGAGTGAAGTGAAGCTT 1202  
 565 GATGCGTATTAAGCGCGCTGTGATTAATAGCGCGCGCGAGAGTGAAGTGAAGCTT 506  
 QY 1203 GGCAGAGTGTGAAGAGCGAAATGCGCAGGCGATGTGTTGTTGAGCTTTTACGCAAT 1262  
 505 GGCAGAGTGTGAAGAGCGAAATGCGCAGGCGATGTGTTGTTGAGCTTTTACGCAAT 446  
 Db 1263 CGTGAATGAAGACGCGATACCGCGCTTGCAATTGTGA 1299  
 445 CGTGAATGAAGACGCGATACCGCGCTTGCAATTGTGA 409

RESULT 9  
 BE520240/c 393 bp mRNA linear EST 19-MAR-2001  
 LOCUS M1B12TM Arabidopsis developing seed Arabidopsis thaliana cDNA  
 DEFINITION clone M1B12 5', mRNA sequence.  
 ACCESSION BE520240  
 VERSION BE520240.1 GI:9778242  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 393)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
 Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing  
 seeds. The metabolic pathway from carbohydrrates to seed oil  
 plant Physiol. 124 (4), 1582-1594 (2000)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 MEDLINE  
 PUBMED  
 11115876  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI  
 48824, USA  
 Tel: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis



Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

## FEATURES

**SOURCE**

```

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="M1B12"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E. coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-,
Site_1: EcoRI; Site_2: XhoI"

```

**ORIGIN**

Query Match	29.7%	Score 386.2;	DB 2;	Length 393;
Best Local Similarity	98.7%;	Pred. No. 8.6e-110;		
Matches 388; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

OY	319	ATTATTGCTGATGTTGACGAGCGGTACCCGTAAACAGCGCAAGCCTTGGCGCGGGCGTG	378
Db	393	ATTATTGCTGATGTTGACGAGCGTACCCGTAAACAGCGCAAGCCTTGGCGCGGGAGTG	334
OY	379	GCACCTGACTGTGCATATAACGTACATACCGAGCAGATACGTCCAGTCCCGATCCGTTA	438
Db	333	GCACCTGACTGTGCATATAACGTACATACCGAGCAGATACGTCCAGTCCCGATCCGTTA	274
OY	439	TTTAAATCCTCTTAAAACTGGCGCTTTGCCAATCTGATTAACGGACGTGACTGACCGCATC	498
Db	273	TTTAAATCCTCTTAAAACTGGCGCTTTGCCAATCTGATTAACGGACGTGACTGACCGCATC	214
OY	499	CTCACACGGGCGAGGAGGCTCAATGCTGACTTACCGGCGATCGGCAAAACGGCGTTTGGC	558
Db	213	CTCACACGGGCGAGGAGGCTCAATGCTGACTTACCGGCGATCGGCAAAACGGCGTTTGGC	154
OY	559	GAACCTGAAACGGGCGCTTAATTTTCGCGCAATCAAACTTGCTGCTTTAAACGTGAGAAACAG	618
Db	153	GAACCTGAAACGGGCGCTTAATTTTCGCGCAATCAAACTTGCTGCTTTAAACGTGAGAAACAG	94
OY	619	GACGAAAGCTGTTCAATTAAACGACGCAATTAACATGGAACCTCAAGTGAAGCCCGCAAT	678
Db	93	GACGAAAGCTGTTCAATTAAACGACGCAATTAACATGGAACCTCAAGTGAAGCCCGCAAT	34
OY	679	GTCCTATTAAACGGGTCGGTAAAGCCTCGCATCA	711
Db	33	GTCCTATTAAAGCCTCGGTAAGCCTCGCATCA	1

Accession	LOCUS	DEFINITION	LOCUS	DEFINITION
AM036132	383 bp	mRNA	linear	EST 18-MAY-2001
AM036132	383 bp	mRNA	linear	EST 18-MAY-2001
EST74508	tomato seed, TAMU lycopersicon esculentum cDNA clone			
CL61E1E3	similar to periplasmic phosphoanhydride phosphohydrolyase precursor, putative, mRNA sequence.			

ACCESSION	AW036132
VERSION	AW036132.1
	GI:5894811

SOURCE *Lycopersicon esculentum* (tomato)

**ORGANISM** *Lycopersicon esculentum*

Eukaryota; Viridiplantae; Strept

Spermatophyta; Magnoliophyta; eu-

1 (bases 1 to 282)  
asterids; Lamiales; Solanales; Scrophulariales; Solanaceae; Solanum

REFERENCE	NUMBERS	White
1 (bases 1 to 383)	110101	8

AULHOKS  
 ALCAIDA, U., MILITE, K  
 HOLT I E. TITON J. E  
 LIANG E.

Bull, L. E., Liang, F., Ople, V., & Niernar

Eugene, C. A.; Bowman, C. D.; McElmurry, Martin, G. B.; Tankasley, S. D. and C.

**TITLE** Generation of ESTs from tomato

JOURNAL  
UNPUBLISHED (1999)

COMMENT: Contact: CUGI

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>

**URES**

```

1..383
/organism="Lycopersicon esculentum"
/mol_type="rRNA"
/cultivar="PA96"
/db_xref="taxon:4081"
/clone="CLE1E23"
/tissue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="XLI-Blue MBF"
/clone_1ib="tomato seed, TAMU"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLEF - Tomato Seed EST library. Directionally cloned
cDNAs inserted into pBluescript SK(-) at 5' end with
EcoRI and 3' end with XhoI site."

```

**ORIGIN**

Query Match	29.5%;	Score 383;	DB 2;	Length 383;
Best Local Similarity	100.0%;	Pred. No. 8.7e-109;		
Matches 383;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy	348		AAAAACAGCGCAACCTCTTCCGCGCGCGGCTGGCACTGACTGTCAATTAACGGTACATAC	407
Db	1		TAAACACGGCGCAACCTTCGCGCGCGGCTGGCACTGACTGTCAATTAACGGTACATAC	60
Oy	408		CCAGCGAGTACGTCCAGTCCCGATCCGTTATTTATCTCTTAAAACTGGCGTTTGCCA	467
Db	61		CCAGCGAGTACGTCCAGTCCCGATCCGTTATTTATCTCTTAAAACTGGCGTTTGCCA	120
Oy	468		ACTGGATTAACGGGAACGTGACTGACCGGATCTCTCAGCAGGGCAGAGGGGTCAATTGCTGA	527
Db	121		ACTGGATTAACCGGAACGTGACTGACCGGATCTCTCAGCAGGGCAGAGGGGTCAATTGCTGA	180
Oy	528		CTTTACCGGGGATCGGCAAAACGGGCGTTTCGGAACTGGAAACGGGCGCTTAATTTTCCGCA	587
Db	181		CTTTACCGGGGATCGGCAAAACGGGCGTTTCGGAACTGGAAACGGGCGCTTAATTTTCCGCA	240
Oy	588		ATCAAACTTGTGCTTTAAACGTGAGAAACAGAACGAAAGCTGTTCAATTAACGACGACATT	647
Db	241		ATCAAACTTGTGCTTTAAACGTGAGAAACAGAACGAAAGCTGTTCAATTAACGACGACATT	300
Oy	648		ACCATCGGAATCAAGGTGAGCGCGGACAAATGTCATTAACCGGTGGGTTAAGCTCGC	707
Db	301		ACCATCGGAATCAAGGTGAGCGCGGACAAATGTCATTAACCGGTGGGTTAAGCTCGC	360
Oy	708		ATCAATGCTGACGAGATATTTC	730
Db	361		ATCAATGCTGACGAGATATTTC	383

RESULT 11  
AW036134/c  
AW036134  
383 bp  
mpva  
linear  
EST 18-MAY-2001

LOCUS	AW036134
DEFINITION	EST274510 tomato seed, <i>Solanum lycopersicon</i>

precursor, putative, mRNA sequen

ACCESSION	AW036134
VERSION	CT:5994813

```

VERSION      AM036134.1  GI:3694613
KEYWORDS     EST

```

NEW WORDS  
ESL.  
Lycopersicon esculentum (tomato)

**ORGANISM** *Lycopodium obscurum* (Linn.)  
**SOURCE**

Opportunism  
Eukaryota: Viridiplantae: Strept

Spermatophyta: Magnoliophyta; eu-

asterids; Lamiales; Solanales; So

REFERENCE 1 (bases 1 to 383)

**AUTHORS**    Alcala, J.,    Vrebalov, J.,    White, R.

Holt, I.E., Liang, F., Upton, J., R

Fujii, C.Y., Bowman, C.L., Nierman

TITLE Martin, G.B., Tankley, S.D. and Giovannoni, J.  
JOURNAL Generation of ESTs from tomato seed tissue  
COMMENT Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source  
1. .383  
Location/Qualifiers

/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CBE1E23"  
/tissue\_type="seeds"  
/dev\_stage="quiescent seed"  
/lab\_host="XLI-Blue MRF"  
/clone\_lib="tomato seed, TAMU"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; CBE - Tomato Seed EST Library. Directionally cloned cDNAs inserted into pBluescript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

## ORIGIN

Query Match 29.5%; Score 383; DB 2; Length 383;  
Best Local Similarity 100.0%; Pred. No. 8.7e-109;  
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 TAAACAGGGAAGCCTTCGCGCGGCTGCACTGATGCAATACGATACATAC 407  
DB 383 TAAACAGGGAAGCCTTCGCGCGGCTGCACTGATGCAATACGATACATAC 324  
QY 408 CCAGCGAGATACGTCGATCCGATCCGATTTAAATCTCTTAAATCTGCGCTTGGCA 467  
DB 323 CCAGCGAGATACGTCGATCCGATCCGATTTAAATCTCTTAAATCTGCGCTTGGCA 264  
QY 468 ACTGATTAACGGAACGTCGATCGGATCTCAGACGGGCGAGAGGTCATTTGCTGA 537  
DB 263 ACTGATTAACGGAACGTCGATCGGATCTCAGACGGGCGAGAGGTCATTTGCTGA 204  
QY 528 CTTACCGGCGATCGGAAACGGGCTTGGCAACTGGAAGGGTCTTAATTTTCGCA 587  
DB 203 CTTACCGGCGATCGGAAACGGGCTTGGCAACTGGAAGGGTCTTAATTTTCGCA 144  
QY 588 ATCAAACTTGCTTAAACGTGGAACAGGAGCAAGCTGTCTTAAACGAGGCAAT 647  
DB 143 ATCAAACTTGCTTAAACGTGGAACAGGAGCAAGCTGTCTTAAACGAGGCAAT 84  
QY 648 ACCATCGGAATCAAGGTGAGCGCGCAATGTCTCATTTAACGGGTGCGTAAGCTTCGC 707  
DB 83 ACCATCGGAATCAAGGTGAGCGCGCAATGTCTCATTTAACGGGTGCGTAAGCTTCGC 24  
QY 708 ATCAATGCTGACGAGATATTTTC 730  
DB 23 ATCAATGCTGACGAGATATTTTC 1

RESULT 12  
LOCUS BU713769 531 bp mRNA linear EST 23-OCT-2003  
DEFINITION BU713769.1 GI:28321125

ACCESSION BU713769.1 GI:28321125  
VERSION BU713769.1  
KEYWORDS Schistosoma japonicum  
SOURCE Schistosoma japonicum  
ORGANISM Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.

## REFERENCE

1 (bases 1 to 531)  
AUTHORS Hu, W., Yan, O., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R., Wang, Z. J., Kong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J., Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J., McManus, D. P., Xue, C. L., Feng, Z., Chen, Z. and Han, Z. G.  
Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource  
Nat. Genet. 35 (2), 139-147 (2003)  
JOURNAL MEDLINE  
PUBMED 12879925  
12873349  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex. 45)  
Fax: 86-21-50801922  
Email: [hanzg@hgc.sh.cn](mailto:hanzg@hgc.sh.cn)

## FEATURES

source

1. .531  
Location/Qualifiers

/organism="Schistosoma japonicum"  
/mol\_type="mRNA"  
/strain="Chinese (Auhui) strain"  
/db\_xref="taxon:6182"  
/sex="Male and female"  
/tissue\_type="whole body"  
/dev\_stage="Adult worms"  
/lab\_host="Mouse and rabbit"  
/clone\_lib="Adult SJC 7/94"  
/note="Vector: Lambda ZAP-II XR.; Site\_1: EcoR I; Site\_2: XhoI I; Several hundred adult Schistosoma japonicum (Auhui, P. R. China, strain, of mixed sex, were perfused from the mesenteries of experimentally infected mice and rabbits at the Queensland Institute of Medical Research, Brisbane, Australia (QIMR), and stored for several months in liquid nitrogen. Subsequently, mRNA was isolated at the QIMR from lysates of these worms by oligo dT chromatography, using a kit from Pharmacia. The mRNA was then shipped to Clontech, Palo Alto, CA, USA, who constructed a cDNA library. First strand synthesis was primed with an oligo-dT-XhoI-primer and synthesized using M-MuLV reverse transcriptase. Second strand synthesis was accomplished with RNase H and T4 DNA polymerase. The double stranded cDNA was ligated to EcoRI linkers, digested with EcoRI and XhoI, and ligated into the phagemid vector lambda ZAP II XR. After construction of this directional library by Clontech, it was returned to the QIMR. During analysis of the library at the QIMR, we have found that a small percentage, 28 to 3%, of the clones contain inserts that appear to be highly homologous to sequences from salmonid fishes, as determined by homology comparisons using BLAST and by Southern hybridization analysis to genomic DNA from salmon (Sigma Chemical Co., St. Louis, MO) under stringent washing conditions. The remainder of the clones appear to contain S. japonicum sequences."

## ORIGIN

Query Match 27.2%; Score 353.4; DB 5; Length 531;  
Best Local Similarity 88.6%; Pred. No. 2e-99; 51; Indels 4; Gaps 4;  
Matches 428; Conservative 0; Mismatches 51;

QY 670 GCCGCAATGCTCTAATTAAACCGGTGCGTAACTCGATCAATGCTGACGAGATATTT 729  
DB 1 GCCGCAATGCTCTAATTAAACCGGTGCGTAACTCGATCAATGCTGACGAGATGCGG 60  
QY 730 CTCCTGCAACAAGCAGGGAATGCCGAGCGGGGTGGGAAGATCACCGATTACAC 789  
DB 61 CTCCTGCAACAAGCAGGGAATGCCGAGCGGGGTGGGAAGATCACCGATTACAC 120  
QY 790 CAGTGAACACCTTGTAAGTTTGCATACGGGCAATTTATTTGTAACAAGCAGCA 849  
DB 121 CAGTGAACACCTTGTAAGTTTGCATACGGGCAATTTATTTGTAACAAGCAGCA 180

Db	351	TCCTTCACGAGGGCAGGA - GGTCAATTGCTGATTTACCGGGGATCGGCAAAACGGGCGCTTTC	293
Qy	557	GCGAAGCTGGAACGGGTCCTAATTTTCCGCAATAAACCCTGTGCTTAAACGTGAGAAC	616
Db	292	GCGAAGCTGGAACGGGTCG - TTAATTTTCCGCAATAAACTGTGCTTAAACGTGAGAAC	234
Qy	617	AGGACGAAAGCTGTTCATTAAACGACGACATTTACCATCGGAACTCAAGTGAAGCCCGACA	676
Db	233	AGGACGAAAGCTGTTCATTAAACGACGACATTTACCATCGGAACTCAAGTGAAGCCCGACA	174
Qy	677	ATGTCCTATTAAACCGGTGGGCTAGCCTCGGATCAATGCTGACGAGATATTTCTCTGCG	736
Db	173	ATGTCCTATTAAACCGGTGGGCTAGCCTCGGATCAATGCTGACGAGATATTTTTCCTGCG	114
Qy	737	AACAAAGCAAGGAGATCCCGGAGCCGGGGTGGGAAAGGATCACCGATTCAACCACTGGA	796
Db	113	AACAAAGCAAGGAGATCCCGGAGCCGGGGTGGGAAAGGATCACCGATTCAACCACTGGA	54
Qy	797	ACACCTTGCTAAGTTTGATATACCGCGCAATTTTATTTTGCTACAAGCAGCGCA	849
Db	53	ACACCTTGCTAAGTTTGATATACCGCGCAATTTTATTTTGCTACAAGCAGCGCA 1	

RESULT	14
LOCUS	CN759004/c
DEFINITION	CN759004 706 bp mRNA linear EST 20-MAY-2004
ACCESSION	ID0AAA24BC04M1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA24BC04
VERSION	5, mRNA sequence.
KEYWORDS	CN759004 GI:47532927
SOURCE	EST.
ORGANISM	Acyrthosiphon pisum (pea aphid)
	Acyrthosiphon pisum
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyigota;
	Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
	Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
REFERENCE	1 (bases 1 to 706)
AUTHORS	Hunter,W., Martinez-Torres,D., Rabhe,Y., Sabater,Munoz,B.,
TITLE	Stern,D., Tagu,D. and Winkler,P.
JOURNAL	An expressed sequence tags database for the pea aphid Acyrthosiphon
COMMENT	pisum Unpublished (2004) Contact: D. Tagu INRA Rennes UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France Tel: +33.2.23.48.51.65 Fax: +33.2.23.48.51.50 Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope PCR Primers FORWARD: CAGGAAACAGCTATGACC Plate: 24 row: C column: 4. Location/Qualifiers 1..706 /organism="Acyrthosiphon pisum" /mol_type="mRNA" /cultivar="developmentstage" /db_xref="taxon:7029" /clone="ID0AAA24BC04" /tissue_type="whole insect" /dev_stage="nymphs and adults (parthenogenetic females)" /lab_host="X1L-Blue" /clone_id="ApMS" /note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Harvesting dates: Soil ; Sowing date: 01/06/1999 ; Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenic stages and both winged and wingless adults were collected for

## ORIGIN

library construction ; experimental condition: long  
photoperiod (16-hr light/8-hr dark at 18 c)"

Query Match 22.7% Score 294.8; DB 7; Length 706;  
Best Local Similarity 99.3%; Pred. No. 6.2e-81;  
Matches 296; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1002 CGCAGTGAAGCTCAACTGAGCGCTTCGCCGTACGCCGATTAACAGCCGCCAGGAGTGA 1061  
DB 706 CGCAGTGAAGCTCAACTGAGCGCTTCGCCGTACGCCGATTAACAGCCGCCAGGAGTGA 647  
QY 1062 ACTGAGTGTGAAGCGTGGCGTGGCTAGCGATTAACAGCCAGTGAATTCAGTTTCCT 1121  
DB 646 ACTGAGTGTGAAGCGTGGCGTGGCTAGCGATTAACAGCCAGTGAATTCAGTTTCCT 587  
QY 1122 GGTCTTCAGACTTTACAGAGATGCGTATTAACAGCCGCTGTCTATTAATACGCCGCC 1181  
DB 586 GGTCTTCAGACTTTACAGAGATGCGTATTAACAGCCGCTGTCTATTAATACGCCGCC 527  
QY 1182 CGAGAGGTGAACCTGAGCGCTGCGAGATGTAAGCGAAATGCGCAGGCGATGTTC 1241  
DB 526 CGAGAGGTGAACCTGAGCGCTGCGAGATGTAAGCGAAATGCGCAGGCGATGTTC 467  
QY 1242 GTTGGCAGGTTTACGCAATCGTGAATGAAGCAGCATACCGCTTGCAATTTGTA 1299  
DB 466 GTTGGCAGGTTTACGCAATCGTGAATGAAGCAGCATACCGCTTGCAATTTGTA 409

## RESULT 15

Bu619443 707 bp mRNA linear EST 01-OCT-2003  
LOCUS Bu619443 NIBB Mochii normalized Xenopus early gastrula library  
DEFINITION Xenopus laevis cDNA clone Xl189d24 5', mRNA sequence.

ACCESSION Bu619443  
VERSION Bu619443.1 GI:37258203

KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus; Xenopus.  
1 (bases 1 to 707)  
Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.

TITLE Expressed genes in X. laevis embryo  
JOURNAL Unpublished (2001)  
COMMENT Contact: Tadasi Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp

The information of this clone is available through the following  
URL.  
http://xenopus.nibb.ac.jp.  
Location/Qualifiers

FEATURES  
source 1..707

/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8335"  
/clone="Xl189d24"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 10.5"  
/clone\_lib="NIBB Mochii normalized Xenopus early gastrula  
library"

## ORIGIN

Query Match 19.7%; Score 256; DB 4; Length 707;  
Best Local Similarity 100.0%; Pred. No. 1e-68;  
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 AATGCCGAGCCGGGGTGGGGAAGATCACCGATTACACGAGTGAACACCTTGCTAAG 809  
DB 1 AATGCCGAGCCGGGGTGGGGAAGATCACCGATTACACGAGTGAACACCTTGCTAAG 60  
QY 810 TTTCATTAACGCGCAATTTATTTGTACACAGCAGCGAGAGTTGCCCGCAGCCGGC 869  
DB 61 TTTCATTAACGCGCAATTTATTTGTACACAGCAGCGAGAGTTGCCCGCAGCCGGC 120  
QY 870 CACCCGTTATTAATTTGATCAAGCAGCGTTGACGCCCATTCACGCAAAAACAGGC 929  
DB 121 CACCCGTTATTAATTTGATCAAGCAGCGTTGACGCCCATTCACGCAAAAACAGGC 180  
QY 930 GTATGAGTGAATTAACCACTTCACTGCTGTTATGCGGACAGATATCTATCTGCG 989  
DB 181 GTATGAGTGAATTAACCACTTCACTGCTGTTATGCGGACAGATATCTATCTGCG 240  
QY 990 AAATCTGGCGGCGCA 1005  
DB 241 AAATCTGGCGGCGCA 256

Search completed: February 10, 2005, 17:33:11  
Job time : 5636.01 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 21:17:58 ; Search time 6258.48 Seconds  
(without alignments)  
10057.285 Million cell updates/sec

Title: US-09-926-375B-7\_COPY\_12653\_13951

Perfect score: 1299  
Sequence: 1 atgaagacatcttaacc.....taccgcctgcagttgtaaa 1299

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 470823 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	3470	6 AX042376	AX042376 Sequence
2	1299	100.0	4060	6 AX042374	AX042374 Sequence
3	1299	100.0	6116	6 AX042375	AX042375 Sequence
4	1299	100.0	6708	6 AX042373	AX042373 Sequence
5	1299	100.0	17732	6 AX042378	AX042378 Sequence
6	1299	100.0	20623	6 AX042372	AX042372 Sequence
7	1294.2	99.6	18882	1 D90735	D90735 Escherichia
8	1294.2	99.6	110000	1 U00096_10	Continuation (11 o
9	1293.2	99.6	1901	1 ECOACPBAA	M58708 Escherichia
10	1293.2	99.6	1901	6 AX356572	AX356572 Sequence
11	1291.6	99.4	1901	1 ECOACPBH	L03371 Escherichia
12	1291.6	99.4	1901	6 AX356570	AX356570 Sequence
13	1291.6	99.4	1901	6 AX356574	AX356574 Sequence
14	1291	99.4	1315	1 AF537219	AF537219 Escherich
15	1290	99.3	1901	1 ECOACPHC	L03372 Escherichia
16	1290	99.3	1901	1 ECOACPHD	L03373 Escherichia
17	1290	99.3	1901	1 ECOACPHE	L03374 Escherichia
18	1290	99.3	1901	1 ECOACPHF	L03375 Escherichia
19	1288.4	99.2	1901	1 ECOACPHA	L03370 Escherichia

20	1288.4	99.2	1901	6 AX356571	AX356571 Sequence
21	1286	99.0	5421	6 AX042377	AX042377 Sequence
22	1282.2	98.7	1323	6 AR108133	AR108133 Sequence
23	1282.2	98.7	1323	6 AR127818	AR127818 Sequence
24	1282.2	98.7	1323	6 AR130956	AR130956 Sequence
25	1282.2	98.7	1323	6 AR492966	AR492966 Sequence
26	1282.2	98.7	1323	6 AX052713	AX052713 Sequence
27	1282.2	98.7	1323	6 AX389666	AX389666 Sequence
28	1282.2	98.7	1323	6 AX356566	AX356566 Sequence
29	1267	97.5	11761	1 AE015127	AE015127 Shigella
30	1265.4	97.4	292504	1 AE016981	AE016981 Shigella
31	1257.4	96.8	1489	1 AY496073	AY496073 Escherich
32	1257.4	96.8	1489	6 AR277753	AR277753 Sequence
33	1246.2	95.9	10029	1 AE005292	AE005292 Escherich
34	1246.2	95.9	327773	1 AP002554	AP002554 Escherich
35	1237.4	95.3	301276	1 AE016758	AE016758 Escherich
36	1170.2	90.1	1272	6 BD074747	BD074747 Novel phy
37	716.6	55.2	212936	2 AC020970	AC020970 Mus muscu
38	447.4	34.4	3279	2 AC020970	AY390262 Citrobact
39	321.8	24.8	114979	2 AC020798	AC020798 Mus muscu
40	305.2	23.5	700	1 ECAKPA	X05471 E.coli appa
41	303.8	23.4	2558	1 AY378096	AY378096 Oesumbac
42	302	23.2	2161	1 AY425982	AY425982 Oesumbac
43	228.6	17.6	10029	1 AE013783	AE013783 Yersinia
44	228.6	17.6	110000	1 BX936398_28	Continuation (29 o
45	228.6	17.6	193050	1 AJ414149	AJ414149 Yersinia

## ALIGNMENTS

RESULT 1	AX042376	AX042376	3470 bp	DNA	linear	PAT 23-NOV-2000
LOCUS	AX042376	Sequence 5 from Patent WO0064247.				
DEFINITION	AX042376	AX042376.1 GI:11340994				
ACCESSION	AX042376					
VERSION	AX042376.1	GI:11340994				
KEYWORDS						
SOURCE						
ORGANISM						
synthetic construct						
other sequences; artificial sequences.						
REFERENCE						
1						
AUTHORS						
TITLE						
JOURNAL						
FORNBERG, C.W., GOLLOVAN, S. and PHILLIPS, J.P.						
Transgenic animals expressing salivary proteins						
Patent: WO 0064247-A 5 02-NOV-2000;						
UNIVERSITY OF GUELPH (CA)						
Location/Qualifiers						
1. .3470						
/organism="synthetic construct"						
/mol_type="unassigned DNA"						
/db_xref="taxon:32630"						
/note="R15/APPA transgene"						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity						
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY 1	ATGAAGCCATCTTAATCCATTCTTTATCTCTTCTGATTCGGTTAACCCGCAATCTGCA	60				
DB 1811	ATGAAGCCATCTTAATCCATTCTTTATCTCTTCTGATTCGGTTAACCCGCAATCTGCA	1870				
QY 61	TTGGCTCAGAGTGAGCCGAGCTGAGAGCTGGAAGTGTGTGATTCAGTTCGTCATGCT	120				
DB 1871	TTGGCTCAGAGTGAGCCGAGCTGAGAGCTGGAAGTGTGTGATTCAGTTCGTCATGCT	1930				
QY 121	GTGGGTCTCCATCAACGAGGCAAGCACTGATGAGGATGTCAACCCGAGAGCATGAGCA	180				
DB 1931	GTGGGTCTCCATCAACGAGGCAAGCACTGATGAGGATGTCAACCCGAGAGCATGAGCA	1990				
QY 181	ACCTGGCCGGTAAACCTGGTGTGCTGACACCGCGCGGTGTGAGTAACTGCTTATCTC	240				
DB 1991	ACCTGGCCGGTAAACCTGGTGTGCTGACACCGCGCGGTGTGAGTAACTGCTTATCTC	2050				

QY 241 GACATTACCAAGCCAGCGTCTGGTAGCCGAGATTGCTGGCGAAAAAGGGCTGCCG 300  
DB 2051 GACATTACCAAGCCAGCGTCTGGTAGCCGAGATTGCTGGCGAAAAAGGGCTGCCG 2110  
QY 301 CAGTCTGGTAGCGGTGGCGATTATTTGCTGATGTGACGAGCGTAAACCGGTAAAAAGGCGGA 360  
DB 2111 CAGTCTGGTAGCGGTGGCGATTATTTGCTGATGTGACGAGCGTAAACCGGTAAAAAGGCGGA 2170  
QY 361 GCGTTGCGCGCGGGCTGGCACCTGACCTGTGCAATTAACCGTACATTAACCGGAGATAGC 420  
DB 2171 GCGTTGCGCGCGGGCTGGCACCTGACCTGTGCAATTAACCGTACATTAACCGGAGATAGC 2230  
QY 421 TCCAGTCCCGGATCCGTTATTTATCTCTTAATAAACTGGGGCTTGGCACTGATTAACGG 480  
DB 2231 TCCAGTCCCGGATCCGTTATTTATCTCTTAATAAACTGGGGCTTGGCACTGATTAACGG 2290  
QY 481 AACGTGACGTAGCGGATCTCTAGCAGGGGAGGAGGATCAATTGCTGACTTTACCGGGCAT 540  
DB 2291 AACGTGACGTAGCGGATCTCTAGCAGGGGAGGAGGATCAATTGCTGACTTTACCGGGCAT 2350  
QY 541 CGGCAAAAGCGGCTTTGCGCAACTGGAACGGGTCTTAATTTTCCGCAATCAAACTTGTGC 600  
DB 2351 CGGCAAAAGCGGCTTTGCGCAACTGGAACGGGTCTTAATTTTCCGCAATCAAACTTGTGC 2410  
QY 601 CTTAAACGTGAGAAACAGGACGAAAGCTTTCTATTACCGCAGCATTTACATCGGAATC 660  
DB 2411 CTTAAACGTGAGAAACAGGACGAAAGCTTTCTATTACCGCAGCATTTACATCGGAATC 2470  
QY 661 AAGGTAGGCGCGACCAATGTCTATTACCGGGGCTTAAGCGCTGCATCAATGTCTAGC 720  
DB 2471 AAGGTAGGCGCGACCAATGTCTATTACCGGGGCTTAAGCGCTGCATCAATGTCTAGC 2530  
QY 721 GAGATATTTCTCTGCAACAAGACAGGAAATGCGGACCGGGGTGGGAAAGATCAC 780  
DB 2531 GAGATATTTCTCTGCAACAAGACAGGAAATGCGGACCGGGGTGGGAAAGATCAC 2590  
QY 781 GATTACACACAGTGAACACTTTGCTAATGTTGCAATACCGCAATTTATTTGCTACA 840  
DB 2591 GATTACACACAGTGAACACTTTGCTAATGTTGCAATACCGCAATTTATTTGCTACA 2650  
QY 841 CGCAGCGCAGAGGTTGCGCGAGCGCGGCAACCGGTTATTTAGATTGATCAAGACAGC 900  
DB 2651 CGCAGCGCAGAGGTTGCGCGAGCGCGGCAACCGGTTATTTAGATTGATCAAGACAGC 2710  
QY 901 TTGACCGCCCATCAACCGCAAAAAGCGGTATGTGTGACATTACCACTTCACTGCTG 960  
DB 2711 TTGACCGCCCATCAACCGCAAAAAGCGGTATGTGTGACATTACCACTTCACTGCTG 2770  
QY 961 TTTATGCGCGGACAGATCTAATCTTGGCAATCTGGGGGCGGCACTGGAGCTCACTG 1020  
DB 2771 TTTATGCGCGGACAGATCTAATCTTGGCAATCTGGGGGCGGCACTGGAGCTCACTG 2830  
QY 1021 AGCGTTCCCGGTGAGCGCGGATTAACGCGCGGAGGTGTGAATGCTGTTTGAACCTG 1080  
DB 2831 AGCGTTCCCGGTGAGCGCGGATTAACGCGCGGAGGTGTGAATGCTGTTTGAACCTG 2890  
QY 1081 CGTGGCTAAGCGATTAACGCGAGTGTGATTCAGGTTTCTTCAAGCTTTACG 1140  
DB 2891 CGTGGCTAAGCGATTAACGCGAGTGTGATTCAGGTTTCTTCAAGCTTTACG 2950  
QY 1141 CAGATGCGGTATTAACCGCGCTGTCTAATTAATACCGCGGAGAGGTGTAACG 1200  
DB 2951 CAGATGCGGTATTAACCGCGCTGTCTAATTAATACCGCGGAGAGGTGTAACG 3010  
QY 1201 CTGGCAGAGTGAAGAGGAATGCGAGGGGATGTTGTTGTCAGGTTTACGAA 1260  
DB 3011 CTGGCAGAGTGAAGAGGAATGCGAGGGGATGTTGTTGTCAGGTTTACGAA 3070  
QY 1261 ATCGTGAATGAACAGCATACCGGCTTGGCAGTTTGA 1299  
DB 3071 ATCGTGAATGAACAGCATACCGGCTTGGCAGTTTGA 3109

RESULT 2  
AA042374 4060 bp DNA linear PAT 23-NOV-2000  
LOCUS AX042374  
DEFINITION Sequence 3 from Patent WO0064247.  
ACCESSION AX042374  
VERSION AX042374.1 GI:11340992  
KEYWORDS  
SOURCE  
ORGANISM  
other sequences; artificial sequences.  
REFERENCE  
1  
AUTHORS Forsberg, C.W., Golovan, S. and Phillips, J.P.  
TITLE Transgenic animals expressing salivary proteins  
JOURNAL Patent: WO 0064247-A 3 02-NOV-2000;  
UNIVERSITY OF GUELPH (CA)  
FEATURES  
source  
1..4060  
location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="R15/AppA + Intron transgene"  
ORIGIN  
Query Match 100.0%; Score 1299; DB 6; Length 4060;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAGCATCTTAATCCATTTTATCTTCTGATTCGTTAACCAGCATCTGCA 60  
DB 1811 ATGAAGCATCTTAATCCATTTTATCTTCTGATTCGTTAACCAGCATCTGCA 1870  
QY 61 TTCGCTCAGTAGAGCCGAGAGCTGAAGCTGGAAGATGGTGTATGTCAATGCTATG 120  
DB 1871 TTCGCTCAGTAGAGCCGAGAGCTGAAGATGGTGTATGTCAATGCTATG 1930  
QY 121 GTCGCTCCCAACCAAGGCCACGCACTGATGACAGATGTCACCCAGACGATGCGCA 180  
DB 1931 GTCGCTCCCAACCAAGGCCACGCACTGATGACAGATGTCACCCAGACGATGCGCA 1990  
QY 1931 GTGCGGTCTCAACCAAGGCCACGCACTGATGACAGATGTCACCCAGACGATGCGCA 2050  
QY 181 ACTGGCGGTTAAACTGGGTGGCTGACACCGCGGGTGGTGAATGCTATCTC 240  
DB 1991 ACTGGCGGTTAAACTGGGTGGCTGACACCGCGGGTGGTGAATGCTATCTC 2050  
QY 241 GACATTACCAAGCCAGCGTCTGGTAGCCGAGATTGCTGGCGAAAAAGGGCTGCCG 300  
DB 2051 GACATTACCAAGCCAGCGTCTGGTAGCCGAGATTGCTGGCGAAAAAGGGCTGCCG 2110  
QY 301 CAGTCTGGTAGCGGTGGCGATTATTTGCTGATGTGACGAGCGTAAACCGGTAAAAAGGCGGA 360  
DB 2111 CAGTCTGGTAGCGGTGGCGATTATTTGCTGATGTGACGAGCGTAAACCGGTAAAAAGGCGGA 2170  
QY 361 GCGTTGCGCGCGGGCTGGCACCTGACCTGTGCAATTAACCGTACATTAACCGGAGATAGC 420  
DB 2171 GCGTTGCGCGCGGGCTGGCACCTGACCTGTGCAATTAACCGTACATTAACCGGAGATAGC 2230  
QY 421 TCCAGTCCCGGATCCGTTATTTATCTCTTAATAAACTGGGGCTTGGCACTGATTAACGG 480  
DB 2231 TCCAGTCCCGGATCCGTTATTTATCTCTTAATAAACTGGGGCTTGGCACTGATTAACGG 2290  
QY 481 AACGTGACGTAGCGGATCTCTAGCAGGGGAGGAGGATCAATTGCTGACTTTACCGGGCAT 540  
DB 2291 AACGTGACGTAGCGGATCTCTAGCAGGGGAGGAGGATCAATTGCTGACTTTACCGGGCAT 2350  
QY 541 CGGCAAAAGCGGCTTTGCGCAACTGGAACGGGTCTTAATTTTCCGCAATCAAACTTGTGC 600  
DB 2351 CGGCAAAAGCGGCTTTGCGCAACTGGAACGGGTCTTAATTTTCCGCAATCAAACTTGTGC 2410  
QY 601 CTTAAACGTGAGAAACAGGACGAAAGCTTTCTATTAAAGCAGGCACTTACATCGGAATC 660  
DB 2411 CTTAAACGTGAGAAACAGGACGAAAGCTTTCTATTAAAGCAGGCACTTACATCGGAATC 2470  
QY 661 AAGGTAGGCGCGACCAATGTCTATTACCGGGTGGTAAAGCTGCTGCATCAATGCTAGC 720

Db	1811	ATGAAAGCCAACTTAAATCCATTTTATCTCTCTGATTCGGGTAAACCCCGCAATCTGCA	1870
Oy	61	TTGCGCTCAAGTGAAGCCCGGAGCTGGAAGCTGGAAAGTGGTGAATTTGCACTGCTCATGGT	120
Db	1871	TTTCCTCAGAGTGAAGCCCGAGAGCTGGAAGCTGGAAAGTGGTGAATTTGCACTGCTCATGGT	1930
Oy	121	GTGGGTCTCTCAACCAAGGCGCAAGCAACTGATGCAAGGATGTCACCCGACAGCGCATGGCCA	180
Db	1931	GTGGGTCTCTCAACCAAGGCGCAAGCAACTGATGCAAGGATGTCACCCGACAGCGCATGGCCA	1990
Oy	181	ACCTGGCCGGTAAAACTGGGTTGGCTGACACCGCGCGTGGTGAAGCTAATCCGCTATCTC	240
Db	1991	ACCTGGCCGGTAAAACTGGGTTGGCTGACACCGCGCGTGGTGAAGCTAATCCGCTATCTC	2050
Oy	241	GGACATTAACCAAGCCGACGCTCTGTGTGACCGAGCGGATTTGCTGGCGAAAAAGGGCTGGCG	300
Db	2051	GGACATTAACCAAGCCGACGCTCTGTGTGACCGAGTTGCTGGCGAAAAAGGGCTGGCG	2110
Oy	301	CAGTCTGGTCAAGGTGGGATTAATGTGATGATGTCGACAGAGGTACCCCGTAAACAGGGGAA	360
Db	2111	CAGTCTGGTCAAGGTGGGATTAATGTGATGATGTCGACAGAGGTACCCCGTAAACAGGGGAA	2170
Oy	361	GCTTTCGCGCGCGGGGCTGGCACTGACTGTGCAATAACGTAATACCCAGGCGATACG	420
Db	2171	GCTTTCGCGCGCGGGGCTGGCACTGACTGTGCAATAACGTAATACCCAGGCGATACG	2230
Oy	421	TCCAGTCCCGATCCGCTTATTTAATCCCTCTAATAAACTGGCGTTTGGCCAACTGGATTAACGG	480
Db	2231	TCCAGTCCCGATCCGCTTATTTAATCCCTCTAATAAACTGGCGTTTGGCCAACTGGATTAACGG	2290
Oy	481	AAAGGATGACAGGAGTCTCTCAGACAGGGGACAGAGGGTCAATATGCTGACTTTACCGGGCAT	540
Db	2291	AAAGGATGACAGGAGTCTCTCAGACAGGGGACAGAGGGTCAATATGCTGACTTTACCGGGCAT	2350
Oy	541	CGGCAAAAGCGGTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC	600
Db	2351	CGGCAAAAGCGGTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC	2410
Oy	601	CTTAAACGTGAGAAACAGAGCGAAAGCTGTTCAATTAACGACAGGATTAACCATCGGAATC	660
Db	2411	CTTAAACGTGAGAAACAGAGCGAAAGCTGTTCAATTAACGACAGGATTAACCATCGGAATC	2470
Oy	661	AAAGTGAAGCGCGCAATATGCTCAATTAACGGTGGGATTAAGCTTCGCAATCTGTACG	720
Db	2471	AAAGTGAAGCGCGCAATATGCTCAATTAACGGTGGGATTAAGCTTCGCAATCTGTACG	2530
Oy	721	GAGATATTTCTCTGCAACAAGCAACAGGGAATGCGGAGCCGAGGTGGGGAGGATCAAC	780
Db	2531	GAGATATTTCTCTGCAACAAGCAACAGGGAATGCGGAGCCGAGGTGGGGAGGATCAAC	2590
Oy	781	GATTCAACAAGTGAACAACCTTGCTAATTTGCAATAACGCGCAATTTATTTGCTACAA	840
Db	2591	GATTCAACAAGTGAACAACCTTGCTAATTTGCAATAACGCGCAATTTATTTGCTACAA	2650
Oy	841	CGCAGCGCAAGGTGGCCCGCAGCCGCGCCACCCCGTATTTAAGTTGATCAAGCAAGCG	900
Db	2651	CGCAGCGCAAGGTGGCCCGCAGCCGCGCCACCCCGTATTTAAGTTGATCAAGCAAGCG	2710
Oy	901	TTGACGCGCCCAATCCACCGCAAAAAACAGGCGATGGTGTGACATTAACCACTTCAGTGTG	960
Db	2711	TTGACGCGCCCAATCCACCGCAAAAAACAGGCGATGGTGTGACATTAACCACTTCAGTGTG	2770
Oy	961	TTTATATGCGCGGACAGATACTAATCTGGCAAACTCTGGCGCGCGCATGGAGCTCACTGG	1020
Db	2771	TTTATATGCGCGGACAGATACTAATCTGGCAAACTCTGGCGCGCGCATGGAGCTCACTGG	2830
Oy	1021	ACGCTTCCCGGTCAACCGGATTAACGCGCGCAGGTGGTGAACGTGGTTTGAACGCTGG	1080
Db	2831	ACGCTTCCCGGTCAACCGGATTAACGCGCGCAGGTGGTGAACGTGGTTTGAACGCTGG	2890
Oy	1081	CGTGGCGTAAAGGATTAACGACGAGATTCAGTTTGGCTGTCTTCAGACTTTTACAG	1140
Db	2891	CGTGGCGTAAAGGATTAACGACGAGATTCAGTTTGGCTGTCTTCAGACTTTTACAG	2950

QY	1141	CAGATGCGGTATTAAGACCGCGCTGCATTAAATACCGCCGCGAGAGGTGAATCTGAC	1200			
Db	2951	CAGATGCGGTATTAAGACCGCGCTGCATTAAATACCGCCGCGAGAGGTGAATCTGAC	3010			
QY	1201	CTGGCAGGATGTGAAGACGAAATGCGCAGGGCAATGTGTTGTTGGCAGGTTTACGCA	1260			
Db	3011	CTGGCAGGATGTGAAGACGAAATGCGCAGGGCAATGTGTTGTTGGCAGGTTTACGCA	3070			
QY	1261	ATCGGATGTGAAGCAGCATACCGCTTGCAAGTTTGTA	1299			
Db	3071	ATCGGATGTGAAGCAGCATACCGCTTGCAAGTTTGTA	3109			
RESULT 4						
LOCUS	AX042373	6708 bp	DNA			
DEFINITION	Sequence 2 from Patent WO0064247.	linear	PAT 23-NOV-2000			
ACCESSION	AX042373					
VERSION	AX042373.1	GI:11340991				
KEYWORDS	synthetic construct					
SOURCE	synthetic construct					
ORGANISM	other sequences; artificial sequences.					
REFERENCE	1					
AUTHORS	Forsberg, C.W., Golovan, S. and Phillips, J.P.					
TITLE	Transgenic animals expressing salivary proteins					
JOURNAL	Patent: WO 0064247-A 2 02-NOV-2000;					
FEATURES	UNIVERSITY OF GUELPH (CA)					
	Location/Qualifiers					
source	1..6708					
	/organism="synthetic construct"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:32630"					
ORIGIN	/note="R15/APPA + intron plasmid with pBLCat3 vector"					
Query Match 100.0%; Score 1299; DB 6; Length 6708;						
Best Local Similarity 100.0%; Pred. No. 0;						
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	ATGAAAGCATCTTATATCCCAATTTTATCTCTTGATTCGGTTAACCCGCAATTCGA	60			
Db	1811	ATGAAAGCATCTTATATCCCAATTTTATCTCTTGATTCGGTTAACCCGCAATTCGA	1870			
QY	61	TTGCGTCAGAGTGAACCCGAGCGTAGAGCTGAAAGTGTGATTTGTCAGTCGATAGT	120			
Db	1871	TTGCGTCAGAGTGAACCCGAGCGTAGAGCTGAAAGTGTGATTTGTCAGTCGATAGT	1930			
QY	121	GTGCGTGTCCCAACCAAGGCCAAGCACTGATGCAAGATGTCAACCCCAAGAGCTGGCA	180			
Db	1931	GTGCGTGTCCCAACCAAGGCCAAGCACTGATGCAAGATGTCAACCCCAAGAGCTGGCA	1990			
QY	181	ACCTGGCGGTAAAACTGGGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTAATTC	240			
Db	1991	ACCTGGCGGTAAAACTGGGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTAATTC	2050			
QY	241	GGACATTACCAACCGCAGCGTGTGTGATGCCAGCGAGATTCTTGGGAAAAAGGGCTGCCG	300			
Db	2051	GGACATTACCAACCGCAGCGTGTGTGATGCCAGCGAGATTCTTGGGAAAAAGGGCTGCCG	2110			
QY	301	CAGCTGTGTGCAAGTGTGCGAATTAATGCTGATGTGCGACGACGTAAACCCGTAACAGGCGAA	360			
Db	2111	CAGCTGTGTGCAAGTGTGCGAATTAATGCTGATGTGCGACGACGTAAACCCGTAACAGGCGAA	2170			
QY	361	GCGTTGCGCGCGGGCTGCGACCTGACTGTGTGCAATTAACCGTACATACCAGGCAAGATACG	420			
Db	2171	GCGTTGCGCGCGGGCTGCGACCTGACTGTGTGCAATTAACCGTACATACCAGGCAAGATACG	2230			
QY	421	TTCAGTCCCGAATCCGTTATTTTAATCCTCTTAAAAACTGGCGTTTGGCAACTGGAATACGCG	480			
Db	2231	TTCAGTCCCGAATCCGTTATTTTAATCCTCTTAAAAACTGGCGTTTGGCAACTGGAATACGCG	2290			

Oy	481	AAAGTGA	CTGA	CCGGAT	CTCTCAG	AGGGG	CGAGG	GGG	CGATTA	TTGGTGA	CTTTAC	CGGGCAT	540					
Db	2291	AAAGTGA	CTGA	CCGGAT	CTCTCAG	AGGGG	CGAGG	GGG	CGATTA	TTGGTGA	CTTTAC	CGGGCAT	2350					
Oy	541	CGGCAAA	CGCGG	TTTCCG	CGA	CTGGA	ACGGG	TGG	CTTAA	TTTTTCCG	CAATCAAA	CTTGTGC	600					
Db	2351	CGGCAAA	CGCGG	TTTCCG	CGA	CTGGA	ACGGG	TGG	CTTAA	TTTTTCCG	CAATCAAA	CTTGTGC	2410					
Oy	601	CTTAAAC	GTGA	AAACGA	GC	GAAG	CTGT	TCATTTAA	CCGAG	CGGTAAG	CCCGCAT	CGGAATC	660					
Db	2411	CTTAAAC	GTGA	AAACGA	GC	GAAG	CTGT	TCATTTAA	CCGAG	CGGTAAG	CCCGCAT	CGGAATC	2470					
Oy	661	AAAGTGA	CCGCCA	CAATGT	CTCATTTAA	CCGAG	CGGTAAG	CCCGCAT	CGGAATCA	AAAGTGA	CC	720						
Db	2471	AAAGTGA	CCGCCA	CAATGT	CTCATTTAA	CCGAG	CGGTAAG	CCCGCAT	CGGAATCA	AAAGTGA	CC	2530						
Oy	721	GAGATAT	TTTCTCC	TCGA	AAACGA	GC	GAAG	ATG	CCGAG	CGGGGTG	GGGAAG	GATCAC	780					
Db	2531	GAGATAT	TTTCTCC	TCGA	AAACGA	GC	GAAG	ATG	CCGAG	CGGGGTG	GGGAAG	GATCAC	2590					
Oy	781	GATTCA	CA	CCAG	GTGA	AAAC	CAC	CTTGTCTA	ATTG	TCATTAAC	GGCA	TTTTTA	ATTGCTCA	840				
Db	2591	GATTCA	CA	CCAG	GTGA	AAAC	CAC	CTTGTCTA	ATTG	TCATTAAC	GGCA	TTTTTA	ATTGCTCA	2650				
Oy	841	CGCAC	CGCCAG	AGGTG	CGCCG	CGAC	GGCGG	CCAC	CCCCGTTATTA	AGTTTGA	TTCATCA	AGAC	900					
Db	2651	CGCAC	CGCCAG	AGGTG	CGCCG	CGAC	GGCGG	CCAC	CCCCGTTATTA	AGTTTGA	TTCATCA	AGAC	2710					
Oy	901	TTGAC	GGCCCCA	TCCAC	CGCA	AAAAA	CAG	GGGAT	GGTGTGA	CATTA	ACCAC	CTTCAG	TCTG	960				
Db	2711	TTGAC	GGCCCCA	TCCAC	CGCA	AAAAA	CAG	GGGAT	GGTGTGA	CATTA	ACCAC	CTTCAG	TCTG	2770				
Oy	961	TTTAT	CGCCCGA	CCAG	TACTAATCTTG	CAAA	TCTGG	CGGCGG	CGCATCTGA	AGCTCA	CTG	1020						
Db	2771	TTTAT	CGCCCGA	CCAG	TACTAATCTTG	CAAA	TCTGG	CGGCGG	CGCATCTGA	AGCTCA	CTG	2830						
Oy	1021	ACGCT	TTCCCG	GTCA	CGCCG	GATTA	ACA	CGCCG	CGAG	TGGTGA	ATCTGTG	TTTGA	ACGCTG	1080				
Db	2831	ACGCT	TTCCCG	GTCA	CGCCG	GATTA	ACA	CGCCG	CGAG	TGGTGA	ATCTGTG	TTTGA	ACGCTG	2890				
Oy	1081	CGTGG	CGCTAA	AGCG	ATTAAC	AG	CCAG	TGGATTCAG	TTTCG	TGGTCTTCC	CAAG	CTTTAC	AG	1140				
Db	2891	CGTGG	CGCTAA	AGCG	ATTAAC	AG	CCAG	TGGATTCAG	TTTCG	TGGTCTTCC	CAAG	CTTTAC	AG	2950				
Oy	1141	CAGAT	GGCGTGA	TAAAA	CGCGCT	GTCTTA	TAATTA	TACCG	CGCCCGG	AGAG	GTGAAA	CTGACC	1200					
Db	2951	CAGAT	GGCGTGA	TAAAA	CGCGCT	GTCTTA	TAATTA	TACCG	CGCCCGG	AGAG	GTGAAA	CTGACC	3010					
Oy	1201	CTGGC	AGAGTGTGA	AGAC	GC	GAATG	CGC	AGG	GCATGTG	TTCTG	TGGCAG	AGTTTAC	GCA	1260				
Db	3011	CTGGC	AGAGTGTGA	AGAC	GC	GAATG	CGC	AGG	GCATGTG	TTCTG	TGGCAG	AGTTTAC	GCA	3070				
Oy	1261	ATCGT	GAATGA	AGAC	GCATTA	CCCG	CTTG	CAAGTTG	TAA	1299								
Db	3071	ATCGT	GAATGA	AGAC	GCATTA	CCCG	CTTG	CAAGTTG	TAA	1299								
RESULT 5																		
LOCUS	AX042378	AX042378	17732	bp	DNA	linear	PAT	23-NOV-2000										
DEFINITION	Sequence 7 from Patent WO0064247.																	
VERSION	AX042378																	
ACCESSION	AX042378.1	GI:11340996																
KEYWORDS																		
SOURCE	.																	



source 1. 12732  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Lama2/Alpa transgene"

Query Match 100.0%; Score 1299; DB 6; Length 17732;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGTTAACCCCGCAATCTGCA 60  
12653 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGTTAACCCCGCAATCTGCA 12712  
61 TTGCTCAGAGTAGAGCCGAGAGTGAAGTGGTGAATGTCACTGCTCATGT 120  
12713 TTGCTCAGAGTAGAGCCGAGAGTGAAGTGGTGAATGTCACTGCTCATGT 12772  
121 GTGCTGCTCAACCAAGGCCAGCACTGATGAGATGTCAACCCAGAGCGATGCCA 180  
12773 GTGCTGCTCAACCAAGGCCAGCACTGATGAGATGTCAACCCAGAGCGATGCCA 12832  
181 ACCTGGCCGGTAAACCTGGTGGCTGACACCGGGGGGTGATGATCGCTATCTC 240  
12833 ACCTGGCCGGTAAACCTGGTGGCTGACACCGGGGGGTGATGATCGCTATCTC 12892  
241 GGAATTATCCAGCCAGCCAGCTGTGTAGCCAGCGATTTGCGGGAAGAGGCTGCCG 300  
12893 GGAATTATCCAGCCAGCCAGCTGTGTAGCCAGCGATTTGCGGGAAGAGGCTGCCG 12952  
301 CAGCTGTGTCAGTCCGATTTATTTGCTGATGTGACAGAGGTACCCGTAACAGGCGGA 360  
12953 CAGCTGTGTCAGTCCGATTTATTTGCTGATGTGACAGAGGTACCCGTAACAGGCGGA 13012  
361 GCGTGTGCGCCGCGGGCTGGACCGTACCTGACATTAACCGATACCCAGAGCATAGC 420  
13013 GCGTGTGCGCCGCGGGCTGGACCGTACCTGACATTAACCGATACCCAGAGCATAGC 13072  
421 TCCAGTCCCGATCCGTTATTTAATCTCTAATAACCTGGCTTGCACACTGGATACGCG 480  
13073 TCCAGTCCCGATCCGTTATTTAATCTCTAATAACCTGGCTTGCACACTGGATACGCG 13132  
481 AACGTGACTGACCGGATCTCAGCAGGCGAGAGGGTCAATTGCTGACTTTACCGGCAT 540  
13133 AACGTGACTGACCGGATCTCAGCAGGCGAGAGGGTCAATTGCTGACTTTACCGGCAT 13192  
541 CGGAAACGGCGCTTTCCGGAACCTGGAAAGGGTCTTAATTTCCGCAATCAACTGTGC 600  
13193 CGGAAACGGCGCTTTCCGGAACCTGGAAAGGGTCTTAATTTCCGCAATCAACTGTGC 13252  
601 CTTAAACGTGAGAAACAGAGAGCTGTTCAATTAACGAGCATTAACATCGGAATC 660  
13253 CTTAAACGTGAGAAACAGAGAGCTGTTCAATTAACGAGCATTAACATCGGAATC 13312  
661 AAGGTAGCGCCGCAATGTCTCATTTAACCGGTGCGTAAAGCTTCGATCATCTGACG 720  
13313 AAGGTAGCGCCGCAATGTCTCATTTAACCGGTGCGTAAAGCTTCGATCATCTGACG 13372  
721 GAGATATTTCTCCCTGCAACAGCAACAGGAAATGCGGAGCCGGGGTGGGGAAGATACC 780  
13373 GAGATATTTCTCCCTGCAACAGCAACAGGAAATGCGGAGCCGGGGTGGGGAAGATACC 13432  
781 GATTCAACACAGTGAACACCTTGTGATTAAGTTGATTAACCGCATTTTATTTCTCAAA 840  
13433 GATTCAACACAGTGAACACCTTGTGATTAAGTTGATTAACCGCATTTTATTTCTCAAA 13492  
841 CGCAGCGCAGAGTTGGCCCGCAGCGCCACCCGTTATTAAGTTGATTAACAGCAGCG 900  
13493 CGCAGCGCAGAGTTGGCCCGCAGCGCCACCCGTTATTAAGTTGATTAACAGCAGCG 13552  
901 TTGACGCCCATTCACCGCAAAACAGGCGTATGTTGATTAACCACTTCAGTGTG 960

13553 TTGACGCCCATTCACCGCAAAACAGGCGTATGTTGATTAACCACTTCAGTGTG 13612  
961 TTGATGCGCGGACACAGATCTAATCTGCAATTCGAGGCGGACATGAGCTCAACTGG 1020  
13613 TTGATGCGCGGACACAGATCTAATCTGCAATTCGAGGCGGACATGAGCTCAACTGG 13672  
1021 ACGCTTCCCGGTACGCGGATTAACAGCCCGCAGGTGTGAACCTGTTTGAACGCTGG 1080  
13673 ACGCTTCCCGGTACGCGGATTAACAGCCCGCAGGTGTGAACCTGTTTGAACGCTGG 13732  
1081 CGTGGCTAAGCGATTAACAGCCAGTGAATCAAGTTTCCGTGCTTCCAGACTTACAG 1140  
13733 CGTGGCTAAGCGATTAACAGCCAGTGAATCAAGTTTCCGTGCTTCCAGACTTACAG 13792  
1141 CAGATCGGTATTAACCGCCGCTGCTAATTAATGCGCCCGGAGAGGTGAACCTGAC 1200  
13793 CAGATCGGTATTAACCGCCGCTGCTAATTAATGCGCCCGGAGAGGTGAACCTGAC 13852  
1201 CTGCGAGATGTTGAAGGCAAAATGCGAGGCGATGTGTTGTCAGAGTTTACGCA 1260  
13853 CTGCGAGATGTTGAAGGCAAAATGCGAGGCGATGTGTTGTCAGAGTTTACGCA 13912  
1261 ATCGTGAATGAAGCAGCATATCCCGCTTGCAGTTGTAA 1299  
13913 ATCGTGAATGAAGCAGCATATCCCGCTTGCAGTTGTAA 13951

RESULT 6  
AX042372 20623 bp DNA linear PAT 23-NOV-2000  
LOCUS Sequence 1 from Patent WO0064247.  
DEFINITION AX042372  
ACCESSION AX042372  
VERSION AX042372.1 GI:11340990  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.

REFERENCE  
1 Forsberg, C.W., Golovan, S. and Philippe, J.P.  
Transgenic animals expressing salivary proteins  
Patent: WO 0064247-A 1 02-NOV-2000;  
JOURNAL UNIVERSITY OF GUELPH (CA)  
LOCATION/Qualifiers

FEATURES  
source 1. 20623  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Lama2/Alpa plasmid"

ORIGIN

Query Match 100.0%; Score 1299; DB 6; Length 20623;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGTTAACCCCGCAATCTGCA 60  
12653 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGTTAACCCCGCAATCTGCA 12712  
61 TTGCTCAGAGTAGAGCCGAGAGTGAAGTGGTGAATGTCACTGCTCATGT 120  
12713 TTGCTCAGAGTAGAGCCGAGAGTGAAGTGGTGAATGTCACTGCTCATGT 12772  
121 GTGCTGCTCAACCAAGGCCAGCACTGATGAGATGTCAACCCAGAGCGATGCCA 180  
12773 GTGCTGCTCAACCAAGGCCAGCACTGATGAGATGTCAACCCAGAGCGATGCCA 12832  
181 ACCTGGCCGGTAAACCTGGTGGCTGACACCGGGGGGTGATGATCGCTATCTC 240  
12833 ACCTGGCCGGTAAACCTGGTGGCTGACACCGGGGGGTGATGATCGCTATCTC 12892  
241 GGAATTATCCAGCCAGCCAGCTGTGTAGCCAGCGATTTGCGGGAAGAGGCTGCCG 300  
12893 GGAATTATCCAGCCAGCCAGCTGTGTAGCCAGCGATTTGCGGGAAGAGGCTGCCG 12952

QY 301 CAGTCTGTGACAGTGCAGATTATATGCTGATGTCAGACGAGCTACCCGTAACCAAGCGGAA 360  
 DB 12955 CAGTCTGTGACAGTGCAGATTATATGCTGATGTCAGACGAGCTACCCGTAACCAAGCGGAA 13012  
 QY 361 GCGTTGCGCGCGCGGCTGCGACCTGACTGTGCAATTAACCGTACATACCAGCGAGATACG 420  
 DB 13013 GCGTTGCGCGCGCGGCTGCGACCTGACTGTGCAATTAACCGTACATACCAGCGAGATACG 13072  
 QY 421 TCCAGTCCCGATCCGTTATTTATATCTCTTAATAAACTGCGGTTTGCCAACTGATTAACGCG 480  
 DB 13073 TCCAGTCCCGATCCGTTATTTATATCTCTTAATAAACTGCGGTTTGCCAACTGATTAACGCG 13132  
 QY 481 AACGTGACTGACGCGATCTCTGACGAGGAGGAGGCTCAATTGCTGACTTTACCGGCGAT 540  
 DB 13133 AACGTGACTGACGCGATCTCTGACGAGGAGGAGGCTCAATTGCTGACTTTACCGGCGAT 13192  
 QY 541 CCGCAAAACGCGCTTTGGCGAATGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTC 600  
 DB 13193 CCGCAAAACGCGCTTTGGCGAATGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTC 13252  
 QY 601 CTTAAACGTGAGAAACAGACGAAAGCTGCTTAATTAACGAGGCAATTAACATCGAACTC 660  
 DB 13253 CTTAAACGTGAGAAACAGACGAAAGCTGCTTAATTAACGAGGCAATTAACATCGAACTC 13312  
 QY 661 AAGGTGAGCGCGCAGCAATGTCTCATTTAACCGGTGCGGTAAAGCTCGCATCAATGCTGAGC 720  
 DB 13313 AAGGTGAGCGCGCAGCAATGTCTCATTTAACCGGTGCGGTAAAGCTCGCATCAATGCTGAGC 13372  
 QY 721 GAATATTTCTCTGCAACAGACAGGGAATGCCGAGCGGGTGGGGAAGGATCACC 780  
 DB 13373 GAATATTTCTCTGCAACAGACAGGGAATGCCGAGCGGGTGGGGAAGGATCACC 13432  
 QY 781 GATTACACACAGTGAACAACCTTGTCTAAAGTTTCATTAACGCGCAATTTTATTTGCTACAA 840  
 DB 13433 GATTACACACAGTGAACAACCTTGTCTAAAGTTTCATTAACGCGCAATTTTATTTGCTACAA 13492  
 QY 841 CGCACGCGCAGAGGTTGCGCGCAGCGCGGCAACCCCGTATTAATTAATTAATTAATTAATTAAT 900  
 DB 13493 CGCACGCGCAGAGGTTGCGCGCAGCGCGGCAACCCCGTATTAATTAATTAATTAATTAATTAAT 960  
 QY 901 TTGACGCGCGCATCCACGCGCAAAAACAGCGGATGATGATGATGATGATGATGATGATGATG 960  
 DB 13553 TTGACGCGCGCATCCACGCGCAAAAACAGCGGATGATGATGATGATGATGATGATGATGATG 13612  
 QY 961 TTTATCCCGGACACGATTAATCTGCGCAATCTCGGCGCGCATGAGCTCAACTGCG 1020  
 DB 13613 TTTATCCCGGACACGATTAATCTGCGCAATCTCGGCGCGCATGAGCTCAACTGCG 13672  
 QY 1021 ACGCTTCCCGGTCAAGCGGATTAACAAGCGCGCGAGTGTGAACCTGTTGAACGCTGCG 1080  
 DB 13673 ACGCTTCCCGGTCAAGCGGATTAACAAGCGCGCGAGTGTGAACCTGTTGAACGCTGCG 13732  
 QY 1081 CGTGGGCTGAAGCATTAACAAGCGCATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 DB 13733 CGTGGGCTGAAGCATTAACAAGCGCATGATGATGATGATGATGATGATGATGATGATGATG 13792  
 QY 1141 CAGATGCGGTGAATAACGCGCTGTCAATTAATACGCGCGCGAGAGGTGAACCTGAC 1200  
 DB 13793 CAGATGCGGTGAATAACGCGCTGTCAATTAATACGCGCGCGAGAGGTGAACCTGAC 13852  
 QY 1201 CTGGCAGAGATGTAAGAGCGAAATGCGCAGGAGGAGTGTGTTGTTGAGAGGTTTAAACGAA 1260  
 DB 13853 CTGGCAGAGATGTAAGAGCGAAATGCGCAGGAGGAGTGTGTTGTTGAGAGGTTTAAACGAA 13912  
 QY 1261 ATCGGATGAAGCAGCATACCGGCTTGCACTTTGTA 1299  
 DB 13913 ATCGGATGAAGCAGCATACCGGCTTGCACTTTGTA 13951

RESULT 7  
 D90735  
 LOCUS D90735 18882 bp DNA linear BCT 25-DEC-2002

DEFINITION *Escherichia coli* K12 genomic DNA. (22.3 - 22.7 min).  
 ACCESSION D90735 AB001340  
 VERSION D90735.1 GI:4662539  
 KEYWORDS Complete and shotgun sequencing; hyaB; hyaC; hyaD; hyaE; hyaF; appC; cyxB; cldA; appB; cyxB; cldB; appA; amsl; inaA; inaB; cspF; cspB; sfa; yccL; yccM  
 SOURCE *Escherichia coli* K12  
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Escherichia*  
 REFERENCE 1 Oshima, T., Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kasahiro, K., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y., Oshima, T., Nishio, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T. A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map DNA Res. 3 (3), 137-155 (1996)  
 JOURNAL 97061202  
 MEDLINE 8905232  
 PUBMED 8905232  
 REFERENCES 2 Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kasahiro, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.  
 TITLE The systematic sequencing of the *Escherichia coli* genome in Japan  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 18882)  
 AUTHORS Mori, H.  
 JOURNAL Direct Submission  
 TITLE Submitted (23-JUL-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5 Takayama, Ikoma, Nara 630-01, Japan  
 JOURNAL (E-mail: hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)  
 COMMENT On Dec 24, 1998 this sequence version replaced gi:1651473.  
 LABORATORY INFORMATION: Collaboration Information:  
 PROJECT: The Japan E. coli genome DNA sequencing project  
 GROUP: The Japan E. coli genome DNA sequencing group  
 MEMBERS: (1995.4 - 1996.3)  
 Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kasahiro, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.  
 HEADED BY:  
 Name: Takashi Horiuchi  
 Address: National Institute of Basic Biology, Okazaki, 444, Japan  
 E-mail: kishori@nibb.ac.jp  
 INFORMATION OPERATOR:  
 Name: Hirotsada Mori  
 Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan  
 E-mail: hmori@gtc.aist-nara.ac.jp  
 URL:  
 The Japan E. coli genome database  
 http://dsb3.aist-nara.ac.jp.  
 Location/Qualifiers  
 1.18882  
 /organism="Escherichia coli K12"  
 /mol\_type="genomic DNA"  
 /strain="K12"  
 /db\_xref="taxon:83333"

## FEATURES

source

/map="22.3-22.7 min"  
/clone="Kohara clone #225"  
/note="Nucleotide position 1034630-1053511 from the  
initiation site of Thra (0 min). This clone is from  
Kohara lambda miniset library."  
1. .966  
/gene="hyaB"  
/gene="hyaB"  
<1. .966  
/gene="hyaB"  
/note="ORF ID:0224#6  
similar to SwissProt Accession Number P19927"  
/codon\_start=1  
/transl\_table=11  
/product="Hydrogenase-1 large chain (EC 1.18.99.1) (nife  
hydrogenase) (membrane-bound hydrogenase 1 large subunit)  
(hyd1)."  
/protein\_id="BAA35738.1"  
/db\_xref="GI:4062540"  
/translation="PDALAGQFNKPKSEIGTGLSDKCVLSYGAFPDIANDGEKSL  
MEGAVINDEFNNVLPVDLVDPOQVEFVDHARYRPNDQVGRHPFDGITDPWYRNGD  
VKGSDNIQOLNEOERYSWIKAPRMGNAMVEGPLARTYLLAHKGDALVESVDRMS  
ALNLPLSIGISTGRIICRAHEAQMAGKLOYPFDKLMTNLKNGNLATISTEKMBPAT  
WTEPCGVGTETAPRGALGHMAAIRDGKIDLYOCVPTTNWASPRDPKQIGAYEAL  
MTKMAIPEQPLEILTLHSFDPCLACSTHVLGDSDSELISQVR"  
985. .1692  
/gene="hyaC"  
/gene="hyaC"  
985. .1692  
/note="ORF ID:0225#1  
similar to PIR Accession Number JV0074"  
/codon\_start=1  
/transl\_table=11  
/product="HyaC protein"  
/protein\_id="BAA35739.1"  
/db\_xref="GI:1651475"  
/translation="MOOKSDNVSHVREAPVRIHMLTVLCMAVLMVWYGFYKPLP  
SVSGEATLTYMGTRILHPSAGNVTFTVLMKITYAFVGNRRISRELFITPVKSKM  
QGVWEIRMYLFLAKRPSADIGHNPIAQAMFGFVSMITGFALYSHEYOYALF  
APRYVVEFPTWGTGNSMDIHSWRLGMVLIGAFVIGHVYMALEDIMSDVTSTW  
NGYSHKFKISNKERS"  
1689. .2276  
/gene="hyaD"  
/gene="hyaD"  
1689. .2276  
/note="ORF ID:0225#2  
similar to SwissProt Accession Number P19930"  
/codon\_start=1  
/transl\_table=11  
/product="Hydrogenase-1 operon protein hyaD."  
/protein\_id="BAA35740.1"  
/db\_xref="GI:1651476"  
/translation="MSEORVVMGLGNLMADEGFAVARELYAHYHMEVEYIVDG  
VTGQINLIGVESASHLLILDIDYGLRPTLRYSAGERI PAVLSAKMSLHONSFE  
VLAADIRGHPAHIALVGLQPMADYGGSLSEIALBQLPAAEQALLAQAANGIYP  
QPAESRCLNVDCLSMENYEGVRLQYRMTOEBG"  
2273. .2671  
/gene="hyaB"  
/gene="hyaB"  
2273. .2671  
/note="ORF ID:0225#3  
similar to SwissProt Accession Number P19931"  
/codon\_start=1  
/transl\_table=11  
/product="Hydrogenase-1 operon protein hyaB."  
/protein\_id="BAA35741.1"  
/db\_xref="GI:1651477"  
/translation="MSNDTPFDALMQMLARGTTPVSESRIDMLTQAPDGVLLSD  
PKRTPEVSDNPEVGMGLREPPDYTWQALADLEQSAIGDRFGVFPPATLVETGAN  
YRGVINGIHPAEELINLMRGIVPEQOERAS"  
2668. .3525  
/gene="hyaF"  
/gene="hyaF"  
2668. .3525  
/gene="hyaF"  
/gene="hyaF"

/note="ORF ID:0225#4  
similar to SwissProt Accession Number P19932"  
/codon\_start=1  
/transl\_table=11  
/product="Hydrogenase-1 operon protein hyaF."  
/protein\_id="BAA35742.1"  
/db\_xref="GI:1651478"  
/translation="MSEPFHLLSPGTQPNDSFSNPLPITCOVNDSPMALEQCA  
HSPQVYALINEIOHOLSEKQPRIGEVLAVDLNLNADDDHPIINTLLGEGEVRIQOA  
DSESEIQAIFCGLMVRKRGKLEKLEKAGCAPLALMQAATONLLPTSLRPP  
IDGLNGLPLAHELAAVNPDAQPHSINTQLPTEADRLFLSLCGGENTQIRITG  
YGESYINATGLRHVHLRCTDTLKGPLLESYBICPIPEVVALAPEDVDSAPRLSEVC  
OMLAEPAP"  
3659. .5203  
/gene="appC"  
/note="synonyms: cbaA, cyxA"  
/gene="appC"  
3659. .5203  
/note="synonyms: cbaA, cyxA"  
/gene="appC"  
/note="ORF ID:0225#5  
similar to SwissProt Accession Number P26459"  
/codon\_start=1  
/transl\_table=11  
/product="Cytochrome bd-ii oxidase subunit i (EC  
1.10.3.-)."  
/protein\_id="BAA35743.1"  
/db\_xref="GI:1651479"  
/translation="MMDVIDLSRWQPALTYLHFLVPTLGLIFLILAMERTVYVTC  
KITIYRDMTRFWGKLFGINFALVATGLTEPFGINWTSFYSNVGIFGAPPLAMEALM  
AFPLESTFVGLFEFGQRLNKYOHLVTVLVAFGSLISLMLINANGMOQPYGAHFD  
IDTLRBMETSFSELVFNVPYOVYFVHTVWAGVYTGAMPINAIISAWLLNGRNVHLR  
SPAIGSPGCTIAIIGTLOLGDSSAYEAVOAPVYKLAAMEGMOETEPAPAPFHVAMP  
QODERNAPALIKTIPALLGITATISLSDKPVGICAKIMAEYTPRLORGMAVLMQETISG  
NEBPHVLAQFRGLSDGGLMLSRYPADMMHTTAQOYQAARGAIPQVAPVNSFRI  
MNGGSLILVMLIALVOTLRKIDQHRVVLKVALMSLPLPVAIEAGFWTEFGROP  
WAQIDILPTYSASHALTTQLAFLIMIVGLYTLFLAELYLMQKYARLRGSPAMQSQ  
PQQOQ"  
5215. .6351  
/gene="appB"  
/note="synonyms: cbaB, cyxB"  
5215. .6351  
/gene="appB"  
/note="ORF ID:0225#6  
similar to SwissProt Accession Number P26458"  
/codon\_start=1  
/transl\_table=11  
/product="Cytochrome bd-ii oxidase subunit ii (EC  
1.10.3.-)."  
/protein\_id="BAA35744.1"  
/db\_xref="GI:1651480"  
/translation="MPDYETLRFTIMWLLIGVILVFMISNGFPMGICLLPIVARND  
ERRIVINSVGAHEGQVWILILGALLFAPMPRYVAAIRSGFYAMILVICSIFPRPL  
AFPRYRKIDARRRKMDGLVLSVPPVFIAGNLLVGPFAFTPOLRVEYLOS  
FMQLLRPFPLCGLLSLGMIYLOGGVWLQKTVGVIHLRSQALTKKAAALLVMLCFLLA

Query Match 99.6%; Score 1294.2; DB 1; Length 18882;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	ATGAAGCATCTTAATCCATTTTATCTCTCTGATTCCTGTTAACCCCGCATATCGCA	60
Dy	6336	ATGAAGCATCTTAATCCATTTTATCTCTCTGATTCCTGTTAACCCCGCATATCGCA	6595
Qy	61	TTGCTCAGAGTGAAGCCGAGCTGAAAGCTGAAAGTGTGTAATGTCAATCGTCATGCT	120
Dy	6596	TTGCTCAGAGTGAAGCCGAGCTGAAAGCTGAAAGTGTGTAATGTCAATCGTCATGCT	6655
Qy	121	GTGCTGCTCCACCAAGGCGACGCACTGATGAGGATGCAACCCGACGACGATGACCA	180
Dy	6656	GTGCTGCTCCACCAAGGCGACGCACTGATGAGGATGCAACCCGACGACGATGACCA	6715
Qy	181	ACCTGCGCGTAAACTGAGTTGGCTGACACCGCGCGTGTGAGCTTAATCGCTATCTC	240
Dy	6716	ACCTGCGCGTAAACTGAGTTGGCTGACACCGCGCGTGTGAGCTTAATCGCTATCTC	6775

QY 241 GACATTACCAAGCCAGCGTCTGGTAGCCAGATTGCTGGCCGAAAAAGGGCTGCCG 300  
 DB 6776 GACATTACCAAGCCAGCGTCTGGTAGCCAGATTGCTGGCCGAAAAAGGGCTGCCG 6835  
 QY 301 CAGTCGTGAGAGTGGCGATTTATGCTGATGTCAGACGATGATCCGTTAAACAGAGCGAA 360  
 DB 6836 CAGTCGTGAGAGTGGCGATTTATGCTGATGTCAGACGATGATCCGTTAAACAGAGCGAA 6895  
 QY 361 GCTTGGCCGCGGGGTGGCACTGATGTCATTAACGTTACATACCCAGGCAATAG 420  
 DB 6896 GCTTGGCCGCGGGGTGGCACTGATGTCATTAACGTTACATACCCAGGCAATAG 6955  
 QY 421 TCCAGTCCCGATCCGTTATTTATCTCTTAATAAACTGGGTTTGGCAATGATTAACGG 480  
 DB 6956 TCCAGTCCCGATCCGTTATTTATCTCTTAATAAACTGGGTTTGGCAATGATTAACGG 7015  
 QY 481 AACGTGACGAGCGATCTCAGCAGAGGAGAGGATGATGCTGATGCTTACCGGGCAT 540  
 DB 7016 AACGTGACGAGCGATCTCAGCAGAGGAGAGGATGATGCTGATGCTTACCGGGCAT 7075  
 QY 541 CGGCAACGCGGCTTTGCGCAACTGAAACGGGTCTTAATTTTCCGATCAAACTTGTGC 600  
 DB 7076 CGGCAACGCGGCTTTGCGCAACTGAAACGGGTCTTAATTTTCCGATCAAACTTGTGC 7135  
 QY 601 CTTAAACGTGAAGAAACAGACGAAAGCTGTCATTAAACGAGGATTAACATCGGAATC 660  
 DB 7136 CTTAAACGTGAAGAAACAGACGAAAGCTGTCATTAAACGAGGATTAACATCGGAATC 7195  
 QY 661 AAGGTGAGCGCGCAATGCTCATTTAACCGGTGCGGTAAAGCTCCGATCAATGCTGAG 720  
 DB 7196 AAGGTGAGCGCGCAATGCTCATTTAACCGGTGCGGTAAAGCTCCGATCAATGCTGAG 7255  
 QY 721 GAGATATTTCTCTGCAACAGCAACAGGATCCGAGCCGGGGTGGGGAAGATCAC 780  
 DB 7256 GAGATATTTCTCTGCAACAGCAACAGGATCCGAGCCGGGGTGGGGAAGATCAC 7315  
 QY 781 GATTCAACAGAGGAAACACTTGTGTAAGTTGATTAAGCGCAATTTTATTTGTACAA 840  
 DB 7316 GATTCAACAGAGGAAACACTTGTGTAAGTTGATTAAGCGCAATTTTATTTGTACAA 7375  
 QY 841 CGCAGCGCAGAGGTGGCCGAGCCGCGCCACCCGTTATTAATTTGATTTGATCAAGACG 900  
 DB 7376 CGCAGCGCAGAGGTGGCCGAGCCGCGCCACCCGTTATTAATTTGATTTGATCAAGACG 7435  
 QY 901 TTGAACGCCCATTCACCGCAAAAACAGCGTATGTTGATCAATTAACCACTTCACTGCTG 960  
 DB 7436 TTGAACGCCCATTCACCGCAAAAACAGCGTATGTTGATCAATTAACCACTTCACTGCTG 7495  
 QY 961 TTTATGCGCGGAAACGATTAATCTGCAAAATCTGGCGGCGCATGAGCTCAACTG 1020  
 DB 7496 TTTATGCGCGGAAACGATTAATCTGCAAAATCTGGCGGCGCATGAGCTCAACTG 7555  
 QY 1021 ACGCTTCCCGGTACCGCGAATAACGCGCCGCAAGTGTGAATGATGTTGAACGCTGG 1080  
 DB 7556 ACGCTTCCCGGTACCGCGAATAACGCGCCGCAAGTGTGAATGATGTTGAACGCTGG 7615  
 QY 1081 CGTGGCTAAGCAATTAACGCAATGATTAAGTTTCCGTGCTTCCGACCTTAACG 1140  
 DB 7616 CGTGGCTAAGCAATTAACGCAATGATTAAGTTTCCGTGCTTCCGACCTTAACG 7675  
 QY 1141 CAGATGCGGTATTAACGCGCTGTCTATTAATACGCGCCCGGAGAGTGAACCTGACC 1200  
 DB 7676 CAGATGCGGTATTAACGCGCTGTCTATTAATACGCGCCCGGAGAGTGAACCTGACC 7735  
 QY 1201 CTGGCAGATGTGAAGAGCGAAATGCGCAGGGCATGTTGCTGTTGCGAAGTTTACGCA 1260  
 DB 7736 CTGGCAGATGTGAAGAGCGAAATGCGCAGGGCATGTTGCTGTTGCGAAGTTTACGCA 7795  
 QY 1261 ATGCTGAATGAAGACGATTAACCGCTTGCAGTTTGTAA 1299  
 DB 7796 ATGCTGAATGAAGACGATTAACCGCTTGCAGTTTGTAA 7834

RESULT 8  
 U00096.10  
 WPCOMMENT  
 Sequence split into 47 fragments LOCUS U00096 Accession U00096  

Fragment Name	Begin	End
U00096_00	1	110000
U00096_01	100001	210000
U00096_02	200001	310000
U00096_03	300001	410000
U00096_04	400001	510000
U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	100000
U00096_10	100001	110000
U00096_11	110001	121000
U00096_12	120001	131000
U00096_13	130001	141000
U00096_14	140001	151000
U00096_15	150001	161000
U00096_16	160001	171000
U00096_17	170001	181000
U00096_18	180001	191000
U00096_19	190001	201000
U00096_20	200001	211000
U00096_21	210001	221000
U00096_22	220001	231000
U00096_23	230001	241000
U00096_24	240001	251000
U00096_25	250001	261000
U00096_26	260001	271000
U00096_27	270001	281000
U00096_28	280001	291000
U00096_29	290001	301000
U00096_30	300001	311000
U00096_31	310001	321000
U00096_32	320001	331000
U00096_33	330001	341000
U00096_34	340001	351000
U00096_35	350001	361000
U00096_36	360001	371000
U00096_37	370001	381000
U00096_38	380001	391000
U00096_39	390001	401000
U00096_40	400001	411000
U00096_41	410001	421000
U00096_42	420001	431000
U00096_43	430001	441000
U00096_44	440001	451000
U00096_45	450001	461000
U00096_46	460001	4639675

 Continuation (11 of 47) of U00096 from base 1000001 (U00096 Escherichia coli K-12 MG1655)

Query Match 99.6%; Score 1294.2; DB 1; Length 110000;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1296; Mismatches 0; Indels 3; Gaps 0;

QY 1 ATGAAGCATCTTAATCCATTTTATCTCTGATTCGTTAAGCCGCAATCTGA 60  
 DB 39840 ATGAAGCATCTTAATCCATTTTATCTCTGATTCGTTAAGCCGCAATCTGA 39899  
 QY 61 TTGCTCAGAGTAGCCGAGCTGAAGTGAAGTGTGATTTGTGATGCTCATGAT 120  
 DB 39900 TTGCTCAGAGTAGCCGAGCTGAAGTGAAGTGTGATTTGTGATGCTCATGAT 39959  
 QY 121 GTGCGTCTCAACCAAGGCCACGCACTGATGCAAGATGTACCCCAAGCGATGGCA 180  
 DB 39960 GTGCGTCTCAACCAAGGCCACGCACTGATGCAAGATGTACCCCAAGCGATGGCA 40019  
 QY 181 ACTGCGCGGTAAACTGGGTTGGCTGACACCGCGGTGTGAGCTAATGCGCTATCTC 240

Db 40020 ACCGCGCGGTAAGAACTGGGTGGCTGACACCGCGCGGTGATGCTAAATCGCTATCTC 40079  
 QY 241 GGAATTACCAACGCGACGCTGTGTAGCCGACGAGTTGCTGCGAAAAAGGCTGCCG 300  
 Db 40080 GGAATTACCAACGCGACGCTGTGTAGCCGACGAGTTGCTGCGAAAAAGGCTGCCG 40139  
 QY 301 CAGTCTGTCAAGGTCCGATTAATTTGCTGATGTGACGAGGTGACCCGTAAACAGCGAA 360  
 Db 40140 CAGTCTGTCAAGGTCCGATTAATTTGCTGATGTGACGAGGTGACCCGTAAACAGCGAA 40199  
 QY 361 GCCTTGGCGCGCGGGGTGGACCTGACCTGCAATTAACGTAACATACCGAGAGATACG 420  
 Db 40200 GCCTTGGCGCGCGGGGTGGACCTGACCTGCAATTAACGTAACATACCGAGAGATACG 40259  
 QY 421 TCCAGTCCCGATCCGTTATTTAAATCTCTTAAAACTGGCGTTTCCCACTGATTAACGCG 480  
 Db 40260 TCCAGTCCCGATCCGTTATTTAAATCTCTTAAAACTGGCGTTTCCCACTGATTAACGCG 40319  
 QY 481 AAGTGAATGACCGCATCTCTCAGCAGGGGAGGAGGGTCAATTTGCTGACTTTACCGGGCAT 540  
 Db 40320 AAGTGAATGACCGCATCTCTCAGCAGGGGAGGAGGGTCAATTTGCTGACTTTACCGGGCAT 40379  
 QY 541 CGGCAACCGCGGTTTGGCGAACTGAAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC 600  
 Db 40380 CGGCAACCGCGGTTTGGCGAACTGAAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC 40439  
 QY 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACCATCGAACTC 660  
 Db 40440 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACCATCGAACTC 40499  
 QY 661 AAGTGAAGCCCGCAATGTCTCTTAAACGGGTGCGTAAAGCTCCGATCAATGCTGACG 720  
 Db 40500 AAGTGAAGCCCGCAATGTCTCTTAAACGGGTGCGTAAAGCTCCGATCAATGCTGACG 40559  
 QY 721 GAGTATTTTCTCCGCAACAGCAGAGATGCGGAGCCGGGTGGGGAGGATCAAC 780  
 Db 40560 GAGTATTTTCTCCGCAACAGCAGAGATGCGGAGCCGGGTGGGGAGGATCAAC 40619  
 QY 781 GATTCAACCAAGTGAACACCTTGTCTAAGTTGCAATTAACGCGCAATTTTAATTGCTACAA 840  
 Db 40620 GATTCAACCAAGTGAACACCTTGTCTAAGTTGCAATTAACGCGCAATTTTAATTGCTACAA 40679  
 QY 841 CGCACCGCAAGGTGCTCCCGACCGCGGCCACCCCGTTATTAAGTTGATCAACAGCGG 900  
 Db 40680 CGCACCGCAAGGTGCTCCCGACCGCGGCCACCCCGTTATTAAGTTGATCAACAGCGG 40739  
 QY 901 TTGACGCGCCATCCACCGCAAAACAGGCGGATGCTGATTAACCACTTCAGTCTG 960  
 Db 40740 TTGACGCGCCATCCACCGCAAAACAGGCGGATGCTGATTAACCACTTCAGTCTG 40799  
 QY 961 TTATTCGCGGACAGATTAATCTGCAATCTCGCGCGGCACTGAGACTCACTG 1020  
 Db 40800 TTATTCGCGGACAGATTAATCTGCAATCTCGCGCGGCACTGAGACTCACTG 40859  
 QY 1021 ACGCTTCCCGGTGACCGGATTAACCGCGCGGAGTGTGAATCTGTGTTGAACGCTGG 1080  
 Db 40860 ACGCTTCCCGGTGACCGGATTAACCGCGCGGAGTGTGAATCTGTGTTGAACGCTGG 40919  
 QY 1081 CGTGGGCTTAAGCAATTAACAGCAGTGAATCAGGTTTGGCTGCTTCACACTTTACAG 1140  
 Db 40920 CGTGGGCTTAAGCAATTAACAGCAGTGAATCAGGTTTGGCTGCTTCACACTTTACAG 40979  
 QY 1141 CAGATGCGGTAAACCGCGCTGTCAATTAATACCGCGCGGAGAGGTGAACCTGACC 1200  
 Db 40980 CAGATGCGGTAAACCGCGCTGTCAATTAATACCGCGCGGAGAGGTGAACCTGACC 41039  
 QY 1201 CTGGCAGGATGTGAAGAGCGAAATGCGCAGGCGCATGTGTTGGCGAGTTTACGAA 1260  
 Db 41040 CTGGCAGGATGTGAAGAGCGAAATGCGCAGGCGCATGTGTTGGCGAGTTTACGAA 41099  
 QY 1261 ATCGGAATGAAGACGATACCGGCTTGACGTTTGTAA 1299  
 Db 41100 ATCGGAATGAAGACGATACCGGCTTGACGTTTGTAA 41138

RESULT 9  
 ECOAPPAA  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 FEATURES  
 source  
 gene  
 RBS  
 CDS  
 stem\_loop  
 -35\_signal  
 -10\_signal  
 RBS  
 CDS  
 Location/Qualifiers  
 1. 1901  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 3. 1522  
 /gene="appa"  
 3. 8  
 /gene="appa"  
 16. 108  
 /gene="appa"  
 /note="ORF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AA72085.1"  
 /db\_xref="GI:145284"  
 /translation="MMYILMFVGLILMCSLTVLVLPRLKS"  
 118. 149  
 /gene="appa"  
 133. 138  
 /gene="appa"  
 156. 161  
 /gene="appa"  
 172. 177  
 /gene="appa"  
 188. 1486  
 /gene="appa"  
 /standard\_name="acid phosphatase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="periplasmic phosphoanhydride phosphohydrolase"  
 /protein\_id="AA72086.1"  
 /db\_xref="GI:145285"  
 /translation="MKAILIPFLSLIPLTPQSAFAPSBEPLKLESVIVISRRGVRA  
 TKATQLMQDVTPEAMPPTWPKLGMILFPRGGELIAYIGHYGRRLVADGLLAKKCCPS

GOVAILADYDERTRKTRGFARFAPGLAPDCAITVHTQADTSSDPPLFNPKTKYCOLDNA  
NNTDAIILSAGSISADFTGROTAFAREBRVINFPOSUICKREKODSCSLTOMLPS  
ELKVSADNVSILGASVLSASMLTEIFLLQAOQMEPFGMRITDSHOMTLISLHNAOF  
YLLQRTPEVARSRAVPLDLIKTALPHRPPQAYGVLLPTSLVFIAHDNVLNAGG  
ALELWLTGEGPDNTPGGELVFERMRRLSDNSQWIOVSLVFQTLQWRDKTPLSLNT  
PGEVFKLTLTAGESERNNAQOMCSLAGFTQVNEARIIPACSL"  
188..253  
/gene="appa"  
254..1483  
/gene="appa"  
/Product="periplasmic phosphoanhydride phosphohydrolyase"  
299..301  
/gene="appa"  
/Standard\_name="H303A mutant"  
/note="created by site directed mutagenesis"  
/citation=[2]  
299..301  
/gene="appa"  
/Standard\_name="R16A mutant"  
/note="created by site directed mutagenesis"  
/citation=[2]  
299..301  
/gene="appa"  
/Standard\_name="R20A mutant"  
/note="created by site directed mutagenesis"  
/citation=[2]  
299..301  
/gene="appa"  
/Standard\_name="R92A mutant"  
/note="created by site directed mutagenesis"  
/citation=[2]  
302..304  
/gene="appa"  
/Standard\_name="H17N mutant"  
/note="created by site directed mutagenesis"  
/citation=[2]  
1163..1165  
/gene="appa"  
/Standard\_name="D304A mutant"  
/note="created by site directed mutagenesis"  
/citation=[3]  
/phenotype="Asp 304 changed to Ala"  
/replace="gct"  
1163..1165  
/gene="appa"  
/Standard\_name="D304E mutant"  
/note="created by site directed mutagenesis"  
/citation=[3]  
/phenotype="Asp 304 changed to Glu"  
/replace="gag"  
1497..1522  
/gene="appa"  
/note="rho-independent terminator"

ORIGIN  
Query Match 99.6%; Score 1293.2; DB 1; Length 1901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAACCATCTTAATCCATTCTTTATCTCTTCTGATTCCGTTAACCCGCAATCTGCA 60  
188 ATGAAACGATCTTAATCCATTCTTTATCTCTTCTGATTCCGTTAACCCGCAATCTGCA 247  
61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGGTGAATTCAGTCCGATGCT 120  
248 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGGTGAATTCAGTCCGATGCT 307  
121 GTGCTGTCTCAACCAAGCGCAGCACTGATGAGAGTGCACCCGAGACGATGAGCA 180  
308 GTGCTGTCTCAACCAAGCGCAGCACTGATGAGAGTGCACCCGAGACGATGAGCA 367  
181 ACCTGCGCGTAAACTGGTTGGCTGACACCGCGCGTGGTGAAGCTTAATCGCCTATCTC 240

368 ACTGCGCGTAAAACTGGTTGGCTGACACCGCGAGGTGGTGAAGCTTAATCGCCTATCTC 427  
241 GGACATTACCAACCCGACCGCTGCTGATGCCGACGAGATTGCTGGGAAAAAGGCTGCCG 300  
428 GACATTACCAACCGCAGCGCTGCTGATGCCGACGAGATTGCTGGGAAAAAGGCTGCCG 487  
301 CAGTCTGTGTCAGTTCGCGATTATTTGCTGATGTCAGACGCTAACCCGTAACACGGCGAA 360  
488 CAGTCTGTGTCAGTTCGCGATTATTTGCTGATGTCAGACGCTAACCCGTAACACGGCGAA 547  
361 GCTTCGCCGCCCGGCTGGCCTGATCTGTGCAATTAACCGTACATACCCAGGCGATACG 420  
548 GCTTCGCCGCCCGGCTGGCCTGATCTGTGCAATTAACCGTACATACCCAGGCGATACG 607  
421 TCCAGTCCCGATCCGTTATTTATCTCTTAATAAACTGGCGCTTGGCCAACTGATACGCG 480  
608 TCCAGTCCCGATCCGTTATTTATCTCTTAATAAACTGGCGCTTGGCCAACTGATACGCG 667  
481 AACGTAAGTGAACGCGATCTCAGACAGGCGAGAGGTCATTTGCTGACTTTACCGGCGAT 540  
668 AACGTAAGTGAACGCGATCTCAGACAGGCGAGAGGTCATTTGCTGACTTTACCGGCGAT 727  
541 CGGCAAAAGCGGCTTTCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAACTGTGCG 600  
728 CGGCAAAAGCGGCTTTCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAACTGTGCG 787  
601 CTTAAACGTGAAGAACGAGACGAAAGCTTTCATTAAACGAGGCTTAACATTCGAACTC 660  
788 CTTAAACGTGAAGAACGAGACGAAAGCTTTCATTAAACGAGGCTTAACATTCGAACTC 847  
661 AAGTGAAGCGCGCAATGTCTCATTTAACGGTGGGTAAAGCTTCGATCAATGCTGACG 720  
848 AAGTGAAGCGCGCAATGTCTCATTTAACGGTGGGTAAAGCTTCGATCAATGCTGACG 907  
721 GAGATATTTCTCTGCAACAGCAACAGGAATCCCGGACCGGGTGGGAAAGATCACG 780  
908 GAGATATTTCTCTGCAACAGCAACAGGAATCCCGGACCGGGTGGGAAAGATCACG 967  
781 GATTCAACACAGTGAACACCTTGCTAAGTTGCAACGCGCAATTTATTTGCTACAA 840  
968 GATTCAACACAGTGAACACCTTGCTAAGTTGCAACGCGCAATTTATTTGCTACAA 1027  
841 CGCAGCGCAGAGGTTGCCGCGACCGCGCACCCCGTTATTTGATTGATCAAGCAGCG 900  
1028 CGCAGCGCAGAGGTTGCCGCGACCGCGCACCCCGTTATTTGATTGATCAAGCAGCG 1087  
901 TTGACGCGCCCATCCACCGCGAATAAGGGGTATGGTGAACCTTACCACTTCACTGCTG 960  
1088 TTGACGCGCCCATCCACCGCGAATAAGGGGTATGGTGAACCTTACCACTTCACTGCTG 1147  
961 TTATGCGCGGACAGATCTAATCTGCAAACTTCGCGGCGCACTGAGGCTCAACTGG 1020  
1148 TTATGCGCGGACAGATCTAATCTGCAAACTTCGCGGCGCACTGAGGCTCAACTGG 1207  
1021 ACGCTTCCGCGTACAGCGGATTAACAGCGCGCAAGTGTGAACTGTGTTGAAGCTGG 1080  
1208 ACGCTTCCGCGTACAGCGGATTAACAGCGCGCAAGTGTGAACTGTGTTGAAGCTGG 1267  
1081 CGTCCGCTAAAGGATTAACAGCGAGTGAATTCAGGTTTCCGCTGCTTCAGACTTACAG 1140  
1268 CGTCCGCTAAAGGATTAACAGCGAGTGAATTCAGGTTTCCGCTGCTTCAGACTTACAG 1327  
1141 CAGATCGTGAATAAAGCGCGCTGTCAATTAAATAGCCCGCGAGAGGTGAATCTGACC 1200  
1328 CAGATCGTGAATAAAGCGCGCTGTCAATTAAATAGCCCGCGAGAGGTGAATCTGACC 1387  
1201 CTGGCAGATGTGAAGAGCGAAATGCGCAGGCGATGTGTTGTTGGCAGGTTTACGCA 1260  
1388 CTGGCAGATGTGAAGAGCGAAATGCGCAGGCGATGTGTTGTTGGCAGGTTTACGCA 1447  
1261 ATGTGAATGAAGACAGCATACCGGCTTGCAATTTGTA 1299  
1448 ATGTGAATGAAGACAGCATACCGGCTTGCAATTTGTA 1486

RESULT 10  
 AX356572 1901 bp DNA linear PAT 06-FEB-2002  
 LOCUS AX356572  
 DEFINITION Sequence 7 from Patent WO0190333.  
 ACCESSION AX356572  
 VERSION AX356572.1 GI:18621059  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli  
 Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE  
 AUTHORS  
 1 Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and  
 O'Donoghue, E.  
 TITLE Recombinant bacterial phylases and uses thereof  
 JOURNAL Patent: WO 0190333-A 729-NOV-2001;  
 DIVERSA CORPORATION (US)  
 FEATURES  
 source  
 1. 1901  
 /organism="Escherichia coli"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:562"  
 ORIGIN  
 Query Match 99.6%; Score 1293.2; DB 6; Length 1901;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Oy 1 ATGAAGCCATCTTAATCCCATTTTATCTCTTCTGATTCCTGTTAACCCCGCATCTGCA 60  
 Db 188 ATGAAGCCATCTTAATCCCATTTTATCTCTTCTGATTCCTGTTAACCCCGCATCTGCA 247  
 Oy 61 TTGGCTCAGAGTAGCCGAGCTGAGCTGAGAAAGTGGAGATTGTCAGTCGTATGCT 120  
 Db 248 TTGGCTCAGAGTAGCCGAGCTGAGCTGAGAAAGTGGAGATTGTCAGTCGTATGCT 307  
 Oy 121 GTGCGTGTCTCAACCAAGGCCACGCACTGATGAGAGATGTCAACCCGAGACGATGCGCA 180  
 Db 308 GTGCGTGTCTCAACCAAGGCCACGCACTGATGAGAGATGTCAACCCGAGACGATGCGCA 367  
 Oy 181 ACCCTGCGCGTAAACTGGGTTGGCTGACACCGCGCGTGTGAGCTTAATCGCTTATCTC 240  
 Db 368 ACCCTGCGCGTAAACTGGGTTGGCTGACACCGCGCGTGTGAGCTTAATCGCTTATCTC 427  
 Oy 241 GGAATTATACCAAGCGCGCTGTGTAAGCGGATTTGCTGGGAAAAAGGCGTGCCTCG 300  
 Db 428 GGAATTATACCAAGCGCGCTGTGTAAGCGGATTTGCTGGGAAAAAGGCGTGCCTCG 487  
 Oy 301 CAGTCTGTGATGATGCGGATTAATGCTGATGTCAGACGATGACCGGTAAACAGCGGAA 360  
 Db 488 CAGTCTGTGATGATGCGGATTAATGCTGATGTCAGACGATGACCGGTAAACAGCGGAA 547  
 Oy 361 GCCTTGCCTGCGCGGCTGGACCTGACTGTGCAATTAACCGTATACCCGAGCATACG 420  
 Db 548 GCCTTGCCTGCGCGGCTGGACCTGACTGTGCAATTAACCGTATACCCGAGCATACG 607  
 Oy 421 TTCCAGTCCCATCCGTTATTTAATCTCTTAATACTGGCGTTTCCCACTGGATTAAGCG 480  
 Db 608 TTCCAGTCCCATCCGTTATTTAATCTCTTAATACTGGCGTTTCCCACTGGATTAAGCG 667  
 Oy 481 AACGTGATGATGATGCGGATCTCAGACGAGGCAAGAGGTCATATGCTGACTTAAACGGG 540  
 Db 668 AACGTGATGATGCGGATCTCAGACGAGGCAAGAGGTCATATGCTGACTTAAACGGG 727  
 Oy 541 CGGCAAAAGCGGTTTCGGGAACTGGAACGGGTGCTTAATTTTCCGCAATCAACTTGTGC 600  
 Db 728 CGGCAAAAGCGGTTTCGGGAACTGGAACGGGTGCTTAATTTTCCGCAATCAACTTGTGC 787  
 Oy 601 CTTAAAGGTGAGAAACGAGAAAGCGTTCATTAACGAGGCAATTAACGATCGGAATC 660  
 Db 788 CTTAAAGGTGAGAAACGAGAAAGCGTTCATTAACGAGGCAATTAACGATCGGAATC 847

Oy 661 AAGGTGAGCGCGGACCAATGCTCATTAACCGGAGCGGTAAAGCTCGCATCAATGCTGACG 720  
 Db 848 AAGGTGAGCGCGGACCAATGCTCATTAACCGGAGCGGTAAAGCTCGCATCAATGCTGACG 907  
 Oy 721 GAGATATTTCTCCTGCAACAGACAGGATATGCGGACCGGGGTGGGAAAGATCACC 780  
 Db 908 GAGATATTTCTCCTGCAACAGACAGGATATGCGGACCGGGGTGGGAAAGATCACC 967  
 Oy 781 GATTACACACAGTGGAACACCTTGTCTAAGTTTGATTAACGCGCAATTTTATTTGCTACAA 840  
 Db 968 GATTACACACAGTGGAACACCTTGTCTAAGTTTGATTAACGCGCAATTTTATTTGCTACAA 1027  
 Oy 841 CGGACCGCGAGGTTTGGCGGCGCGGCGCACCCCGTTATTTGATTTGATCAAGACAGCG 900  
 Db 1028 CGGACCGCGAGGTTTGGCGGCGCGGCGCACCCCGTTATTTGATTTGATCAAGACAGCG 1087  
 Oy 901 TTGACCGCCCATATCAACCGCAAAACAGGGTATGTTGATCAATTAACCACTTCAGTCTG 960  
 Db 1088 TTGACCGCCCATATCAACCGCAAAACAGGGTATGTTGATCAATTAACCACTTCAGTCTG 1147  
 Oy 961 TTTATGCGCGGACAGATCTAATCTTGCAAAATCTGGGCGGCACTGAGTCAACTGCG 1020  
 Db 1148 TTTATGCGCGGACAGATCTAATCTTGCAAAATCTGGGCGGCACTGAGTCAACTGCG 1207  
 Oy 1021 ACCGTTCCCGGTGACCGGATTAACAGCGCGGCGGAGTGAAGTGGTTTGAACGCTGCG 1080  
 Db 1208 ACCGTTCCCGGTGACCGGATTAACAGCGCGGCGGAGTGAAGTGGTTTGAACGCTGCG 1267  
 Oy 1081 CGTGGCTAAGAGATTAACAGCGGATTAAGTTCAGTTCGCTGCTTCCAGACTTTACAG 1140  
 Db 1268 CGTGGCTAAGAGATTAACAGCGGATTAAGTTCAGTTCGCTGCTTCCAGACTTTACAG 1327  
 Oy 1141 CAGATCGGTATTAACCGCGCTGTCAATTAATACCGCGCGGAGAGGTGAACCTGACC 1200  
 Db 1328 CAGATCGGTATTAACCGCGCTGTCAATTAATACCGCGCGGAGAGGTGAACCTGACC 1387  
 Oy 1201 CTGGCAGATGTTGAAGAGCAATATGGCAGGCGGATGTTGCTGGCAGGTTTACGCA 1260  
 Db 1388 CTGGCAGATGTTGAAGAGCAATATGGCAGGCGGATGTTGCTGGCAGGTTTACGCA 1447  
 Oy 1261 ATCGTGAATGAACAGCATATCCCGCTTGCACTTTGTA 1299  
 Db 1448 ATCGTGAATGAACAGCATATCCCGCTTGCACTTTGTA 1486  
 RESULT 11  
 ECOACPHB 1901 bp DNA linear BCT 04-APR-2002  
 LOCUS ECOACPHB  
 DEFINITION Escherichia coli periplasmic phosphoanhydride phosphohydrolase  
 (apppa) gene, complete cds.  
 ACCESSION I03371  
 VERSION I03371.1 GI:145178  
 KEYWORDS acid phosphatase; periplasmic phosphoanhydride phosphohydrolase.  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE  
 AUTHORS  
 1 (sites)  
 Dassa, J., Marcq, C. and Boquet, P.L.  
 TITLE The complete nucleotide sequence of the Escherichia coli gene apppa  
 reveals significant homology between pH 2.5 acid phosphatase and  
 glucose-1-phosphatase  
 JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)  
 MEDLINE 90368616  
 PUBMED 2168385  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 1901)  
 Oetman, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van  
 Overbeek, R.L.  
 TITLE Overexpression, site-directed mutagenesis, and mechanism of  
 Escherichia coli acid phosphatase  
 JOURNAL J. Biol. Chem. 267 (32), 22830-22836 (1992)  
 MEDLINE 93054596



PUBMED 1429631 Location/Qualifiers  
FEATURES  
source 1..1901  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
gene 188..1486  
/gene="appa"  
188..1486  
CDS  
/gene="appa"  
/standard\_name="acid phosphatase"  
/citation=[1]  
/codon\_start=1  
/transl\_table=11  
/product="periplasmic phosphoanhydride phosphohydrolyase"  
/protein\_id="AAA0003.1"  
/db\_xref="GI:145179"  
/translation="MKAILIPLSLILPLTPQSAFAPQSEPELKLSEVIVSRNGVRAP  
TYATQIMDVTPDAMPVPMVKGMTPRQGLIAYIGHYORORLWADGLAKKGRPS  
GVVAITADYDERTRKTEGAFAAGAPDCAITHTOADSPPPLPNPLKTVCOLDNA  
NVTDAILSRAGSIADFTGHRQTAFLERLVRINPQSNLCLEKQDSCTLOALPS  
ELKVSADNVSLTGAIVSLASMLTEIFLQAOQMPBQWGRITDSQWMTLSLNAQF  
YLQRTPEVARSRATPLIDLITKALTPHPQQAQGVTLPLTSLFAGIDTILAMIG  
ALEIMWTLHGPDNTPGGELVEFRMRRLSDNSOMLOVSLVQTLQOMRDKPLSLNT  
PGEVYKTLTAGCEBERRNAQMCSLAGFTQIVNEARIPAGSL"  
302..304  
variation  
/gene="appa"  
/citation=[1]  
/phenotype="H17N mutant"  
ORIGIN  
Query Match 99.4%; Score 1291.6; DB 1; Length 1901;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
1 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGGTTAACCCCGCAATCTGCA 60  
188 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGGTTAACCCCGCAATCTGCA 247  
61 TTGCTCAGAGTAGCCCGAGCTGAAGCTGGAAGTGTGTGATTCAGTCTCATGGT 120  
248 TTGCTCAGAGTAGCCCGAGCTGAAGCTGGAAGTGTGTGATTCAGTCTCATGGT 307  
121 GTGCTGCTCAGAGTAGCCCGAGCTGAAGCTGGAAGTGTGTGATTCAGTCTCATGGT 180  
308 GTGCTGCTCAGAGTAGCCCGAGCTGAAGCTGGAAGTGTGTGATTCAGTCTCATGGT 367  
181 ACCTGGCCGGTAAACCTGGGTTGAGTGAACACCGCGCGTGTGAGTGAATCGCTATCTC 240  
368 ACCTGGCCGGTAAACCTGGGTTGAGTGAACACCGCGCGTGTGAGTGAATCGCTATCTC 427  
241 GACATTTACCAACGCCAGCTGTGTGAGCCGAGGATTCGTGCGAAGAAAGGCTGCGCG 300  
428 GACATTTACCAACGCCAGCTGTGTGAGCCGAGGATTCGTGCGAAGAAAGGCTGCGCG 487  
301 CAGTCTGATGAGTGTGAGTGAATTCGTGAGTGAAGGATTCGTGAGTGAAGGAGGAA 360  
488 CAGTCTGATGAGTGTGAGTGAATTCGTGAGTGAAGGATTCGTGAGTGAAGGAGGAA 547  
361 GCTTTCGCGCGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420  
548 GCTTTCGCGCGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 607  
421 TCCAGTCCGATCCGTTATTTAATCTCTTAAACCTGGCTTTGCCAATGATTAACGCG 480  
608 TCCAGTCCGATCCGTTATTTAATCTCTTAAACCTGGCTTTGCCAATGATTAACGCG 667  
481 AACGTGATGAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 540  
668 AACGTGATGAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 727  
541 CGGCAACGCGGCTTTCGCGAATGGAACGCGGCTTATTTTCCGCAATCAAACTTGTGC 600

Db 728 CGGCAACGCGGCTTTCGCGAATGGAACGCGGCTTATTTTCCGCAATCAAACTTGTGC 787  
Qy 601 CTTAAACGTGAGAAACAGAGCAAGCGTTTCAATTAACGAGCATTAACCATCGGAATC 660  
Db 788 CTTAAACGTGAGAAACAGAGCAAGCGTTTCAATTAACGAGCATTAACCATCGGAATC 847  
Qy 661 AAGGTGAGCGCGCGCAATGTTCTCATTTAAACCGGTGCGGTAAACCTCGATCAATGCTGACG 720  
Db 848 AAGGTGAGCGCGCGCAATGTTCTCATTTAAACCGGTGCGGTGAGCTCGATCAATGCTGACG 907  
Qy 721 GAGATATTTCTCTGCAACAAAGCAAGGAAATGCCCGAGCGCGGTGGGAAAGATCAAC 780  
Db 908 GAGATATTTCTCTGCAACAAAGCAAGGAAATGCCCGAGCGCGGTGGGAAAGATCAAC 967  
Qy 781 GATTCAACACAGTGAACACCTTGCTAAGTTTGATTAACGGCAATTTATTTGCTACAA 840  
Db 968 GATTCAACACAGTGAACACCTTGCTAAGTTTGATTAACGGCAATTTATTTGCTACAA 1027  
Qy 841 CGCACGCGAGGTTGCGCGGAGCGCGCAACCCCGTTATTAGATTGATCAAGACAGCG 900  
Db 1028 CGCACGCGAGGTTGCGCGGAGCGCGCAACCCCGTTATTAGATTGATCAAGACAGCG 1087  
Qy 901 TTGACGCGCCATCCACCGCAAAAACAGCGCTATGTTGATTAACCATTCCTGCTG 960  
Db 1088 TTGACGCGCCATCCACCGCAAAAACAGCGCTATGTTGATTAACCATTCCTGCTG 1147  
Qy 961 TTTATCGCGCGACAGTACTAATCTGCAATCTCGCGCGGCACTGAGACTCAACTG 1020  
Db 1148 TTTATCGCGCGACAGTACTAATCTGCAATCTCGCGCGGCACTGAGACTCAACTG 1207  
Qy 1021 ACGCTTCCCGGTGACCGGATTAACACGCGCGAGTGTGAATCTGTGTTGAACGCTG 1080  
Db 1208 ACGCTTCCCGGTGACCGGATTAACACGCGCGAGTGTGAATCTGTGTTGAACGCTG 1267  
Qy 1081 CGTGGCTTAAGGATTAACAGCGAGTGAATGAGTTTGGCTGCTTCCAGACTTTACG 1140  
Db 1268 CGTGGCTTAAGGATTAACAGCGAGTGAATGAGTTTGGCTGCTTCCAGACTTTACG 1327  
Qy 1141 CAGATGCGGATTAACAGCGCGCTGCTATTAAATACGCGCGGAGAGTGAATCTGAC 1200  
Db 1328 CAGATGCGGATTAACAGCGCGCTGCTATTAAATACGCGCGGAGAGTGAATCTGAC 1387  
Qy 1201 CTGGCAGATGTAAGAGCGAATGCGAGGAGTGTGTTGTTGAGGAGTTTACGAA 1260  
Db 1388 CTGGCAGATGTAAGAGCGAATGCGAGGAGTGTGTTGTTGAGGAGTTTACGAA 1447  
Qy 1261 ATCGTGAATGAAGCAGCATACCCGCTTGCAGTTTGTAA 1299  
Db 1448 ATCGTGAATGAAGCAGCATACCCGCTTGCAGTTTGTAA 1486  
RESULT 12  
AX356570 1901 bp DNA linear PAT 06-FEB-2002  
LOCUS AX356570  
DEFINITION Sequence 5 from Patent WO0190333.  
ACCESSION AX356570  
VERSION AX356570.1 GI:18621057  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1 Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and  
AUTHORS O'Donoghue, B.  
TITLE Recombinant bacterial phytases and uses thereof  
JOURNAL Patent: WO 0190333-A 5 29-NOV-2001;  
DIVERSA CORPORATION (US)  
FEATURES  
source 1..1901  
/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"



## ORIGIN

Query Match 99.4%; Score 1291.6; DB 6; Length 1901;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGAAAGCATCTTAATCCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATGCA 60  
 Db ATGAAAGCATCTTAATCCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATGCA 247  
 Qy TTGCGCTCAGAGTAGAGCCGAGCTGAAAGTGGTGAATGTCAGTCGATGCT 120  
 Db TTGCGCTCAGAGTAGAGCCGAGCTGAAAGTGGTGAATGTCAGTCGATGCT 307  
 Qy 121 GTGCGTCTCAACCAAGGCAACGCACTGATGCAAGATGTCACCCCAAGCGATGCA 180  
 Db GTGCGTCTCAACCAAGGCAACGCACTGATGCAAGATGTCACCCCAAGCGATGCA 367  
 Qy 181 ACCTGCGCGGTAAATCTGGGTGGTGAACACCGCGGGGTGGTGAATGCTAATCGCTATCTC 240  
 Db ACCTGCGCGGTAAATCTGGGTGGTGAACACCGCGGGGTGGTGAATGCTAATCGCTATCTC 427  
 Qy 241 GGCATTACCAAGCCGAGCGTCTGTAGCCGAGCGATGCTGCGCAAAAAGGCGTCCCG 300  
 Db GGCATTACCAAGCCGAGCGTCTGTAGCCGAGCGATGCTGCGCAAAAAGGCGTCCCG 487  
 Qy 301 CAGTCTGTCAAGTCCGATTAATGCTGATGTCAGACGAGCTACCCGTAACAGCGGAA 360  
 Db CAGTCTGTCAAGTCCGATTAATGCTGATGTCAGACGAGCTACCCGTAACAGCGGAA 547  
 Qy 361 GCGTTGCGCGCGCGGCTGCGCACTGACTGTGCAATACCTAATACCCGAGCAATACG 420  
 Db GCGTTGCGCGCGCGGCTGCGCACTGACTGTGCAATACCTAATACCCGAGATACG 607  
 Qy 548 GCGTTGCGCGCGCGGCTGCGCACTGACTGTGCAATACCTAATACCCGAGCAATACG 607  
 Db 421 TCCAGTCCCGATCGTTATTTATCCCTAATACTGCGCTTGCCTGCAATGCAATACGCG 480  
 Db 608 TCCAGTCCCGATCGTTATTTATCCCTAATACTGCGCTTGCCTGCAATGCAATACGCG 667  
 Qy 481 AACGTGATGACCGCATCTCAGACGAGGCAAGAGGTCATATGCTGACTTTACCGGCGAT 540  
 Db AACGTGATGACCGCATCTCAGACGAGGCAAGAGGTCATATGCTGACTTTACCGGCGAT 727  
 Qy 668 AACGTGATGACCGCATCTCAGACGAGGCAAGAGGTCATATGCTGACTTTACCGGCGAT 727  
 Db 541 CCGCAAAAGCGCTTTCGCGCAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC 600  
 Db 728 CCGCAAAAGCGCTTTCGCGCAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC 787  
 Qy 601 CTTAAACGTGAAGAACGAAAGCTGTTCAATTAAGAGGATTAACCATCGGAATC 660  
 Db CTTAAACGTGAAGAACGAAAGCTGTTCAATTAAGAGGATTAACCATCGGAATC 847  
 Qy 788 CTTAAACGTGAAGAACGAAAGCTGTTCAATTAAGAGGATTAACCATCGGAATC 847  
 Db 661 AAGGTGAGCGCCGACATGCTCTCAATTAACGGTGGTAAAGCTTCGATCAATGCTGACG 720  
 Db 848 AAGGTGAGCGCCGACATGCTCTCAATTAACGGTGGTAAAGCTTCGATCAATGCTGACG 907  
 Qy 721 GAGATATTTCTCTGCAACAGCACAGGAAATGCGGAGCCGGGTGGGGAAGATCAAC 780  
 Db GAGATATTTCTCTGCAACAGCACAGGAAATGCGGAGCCGGGTGGGGAAGATCAAC 967  
 Qy 781 GATTCAACAGAGTGAACACCTTGCTAAGTTGCTAAGCGCAATTTTATTTGCTCAAA 840  
 Db GATTCAACAGAGTGAACACCTTGCTAAGTTGCTAAGCGCAATTTTATTTGCTCAAA 1027  
 Qy 841 CGCAGCGCAGAGGTGGCCGAGCGCGCCACCCGTTATTAAGTTTGAATCAAGACGCG 900  
 Db CGCAGCGCAGAGGTGGCCGAGCGCGCCACCCGTTATTAAGTTTGAATCAAGACGCG 1087  
 Qy 901 TTGACGCGCCCATCCACGCAAAAACAGCGGTATGCTGACATTAACCATTCAGTCTG 960  
 Db TTGACGCGCCCATCCACGCAAAAACAGCGGTATGCTGACATTAACCATTCAGTCTG 1147  
 Qy 961 TTATATGCGCGGACAGATCTAATCTGGCAAAATCTCGGCGGCGCATGAGACTCAACTGG 1020  
 Db TTATATGCGCGGACAGATCTAATCTGGCAAAATCTCGGCGGCGCATGAGACTCAACTGG 1207

Qy 1021 AGCTTCCCGGTCAGCCGATTAACAGCCGAGGTGTAATGCTGTTGAACGCTGG 1080  
 Db 1208 AGCTTCCCGGTCAGCCGATTAACAGCCGAGGTGTAATGCTGTTGAACGCTGG 1267  
 Qy 1081 CGTGGCTAAGGATTAACAGCCAGTGAATGAGTTTCGCTGCTTCAGACTTTACAG 1140  
 Db 1268 CGTGGCTAAGGATTAACAGCCAGTGAATGAGTTTCGCTGCTTCAGACTTTACAG 1327  
 Qy 1141 CAGATCGGTATTAACCGCGCTGCTATTAATAGCCGCCCGGAGAGGTGAACCTGACC 1200  
 Db 1328 CAGATCGGTATTAACCGCGCTGCTATTAATAGCCGCCCGGAGAGGTGAACCTGACC 1387  
 Qy 1201 CTGGCAGATGTAAGAGGAAATGGCGAGGAGATGTTTCGTTGACAGTTTACGCA 1260  
 Db 1388 CTGGCAGATGTAAGAGGAAATGGCGAGGAGATGTTTCGTTGACAGTTTACGCA 1447  
 Qy 1261 ATCGTGAATGAAGACGATACCCGCTTCAGTTGTAA 1299  
 Db 1448 ATCGTGAATGAAGACGATACCCGCTTCAGTTGTAA 1486

RESULT 13  
 AX356574 1901 bp DNA linear PAT 06-FEB-2002  
 LOCUS AX356574  
 DEFINITION Sequence 9 from Patent WO0190333.  
 ACCESSION AX356574  
 VERSION AX356574.1 GI:18621060  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli  
 Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

REFERENCE  
 AUTHORS Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and O'Donoghue, B.  
 TITLE Recombinant bacterial phylases and uses thereof  
 JOURNAL Patent: WO 0190333-A 9 29-NOV-2001;  
 DIVERSA CORPORATION (US)  
 FEATURES  
 source Location/Qualifiers  
 1. 1901  
 /organism="Escherichia coli"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:562"

## ORIGIN

Query Match 99.4%; Score 1291.6; DB 6; Length 1901;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGAAAGCATCTTAATCCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATGCA 60  
 Db 188 ATGAAAGCATCTTAATCCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATGCA 247  
 Qy 61 TTGCGCTCAGAGTAGAGCCGAGCTGAAAGTGGTGAATGTCAGTCGATGCT 120  
 Db 248 TTGCGCTCAGAGTAGAGCCGAGCTGAAAGTGGTGAATGTCAGTCGATGCT 307  
 Qy 121 GTGCGTCTCAACCAAGGCAACGCACTGATGCAAGATGTCACCCCAAGCGATGCA 180  
 Db 308 GTGCGTCTCAACCAAGGCAACGCACTGATGCAAGATGTCACCCCAAGCGATGCA 367  
 Qy 181 ACCTGCGCGGTAAATCTGGGTGGTGAACACCGCGGGGTGGTGAATGCTAATCGCTATCTC 240  
 Db 368 ACCTGCGCGGTAAATCTGGGTGGTGAACACCGCGGGGTGGTGAATGCTAATCGCTATCTC 427  
 Qy 241 GGCATTACCAAGCCGAGCGTCTGTAGCCGAGCGATGCTGCGCAAAAAGGCGTCCCG 300  
 Db 428 GGCATTACCAAGCCGAGCGTCTGTAGCCGAGCGATGCTGCGCAAAAAGGCGTCCCG 487  
 Qy 301 CAGTCTGTCAAGTCCGATTAATGCTGATGTCAGACGAGCTACCCGTAACAGCGGAA 360  
 Db 488 CAGTCTGTCAAGTCCGATTAATGCTGATGTCAGACGAGCTACCCGTAACAGCGGAA 547

Qy	361	GCCTTGGCCGCGCGGCTGGGACCTGACCTGTGAATTAACGTATACATACCAGGACATACG	420		
Db	548	GCCCTTCCGCCCGGGCTGGGACCTGTGCAATTAACGTAACATACCAGGACATACG	607		
Qy	421	TCCAGTCCCGCATCCGTTATTTAAATCTCTTAAAAAAGTGGGCTTGGCAACTGTGAATACGG	480		
Db	608	TCCAGTCCCGCATCCGTTATTTAAATCTCTTAAAAAAGTGGGCTTGGCAACTGTGAATACGG	667		
Qy	481	AACGTGACTGAACGGATCTCAAGAGGGACAGAGGGGTCAATTGCTGACCTTTACCGGGCAT	540		
Db	668	AACGTGACTGAACGGATCTCAAGAGGGACAGAGGGGTCAATTGCTGACCTTTACCGGGCAT	727		
Qy	541	CGGGAACCGCGGTTTGGCGAACTGGAAACGGGAGCTTTAAATTTTCCGAATCAAACTGTGAC	600		
Db	728	CGGGAACCGCGGTTTGGCGAACTGGAAACGGGAGCTTTAAATTTTCCGAATCAAACTGTGAC	787		
Qy	601	CTTAAACGTGGAAGAACAGACGAAGAGCTGTTCATTTACCGAGGCATTTACATCGGAATC	660		
Db	788	CTTAAACGTGGAAGAACAGACGAAGAGCTGTTCATTTACCGAGGCATTTACATCGGAATC	847		
Qy	661	AAGGTGACGCGCGAATATGTCTCATTTACCGGTGCGGTAAAGCTTCGCATCAATAGCGAG	720		
Db	848	AAGGTGACGCGCGAATATGTCTCATTTACCGGTGCGGTAAAGCTTCGCATCAATAGCGAG	907		
Qy	721	GAGATATTTCTCCGCAACAAAGCACAAGGAAATGCGGAGCGCGGAGTGGGGAAGATCAAC	780		
Db	908	GAGATATTTCTCTCTGCAACAAAGCACAAGGAAATGCGGAGCGCGGAGTGGGGAAGATCAAC	967		
Qy	781	GATTCAACACAGTGGAAACACCTTGCTAAAGTTGCAATACCGGCATTTATTTTGCTACAA	840		
Db	968	GATTCAACACAGTGGAAACACCTTGCTAAAGTTGCAATACCGGCATTTATTTTGCTACAA	1027		
Qy	841	CGCAGCGCCAGAGGTGGCCCGGACGCGGCCACCCGCTATTGAATTGATCAAGACAGCG	900		
Db	1028	CGCAGCGCCAGAGGTGGCCCGGACGCGGCCACCCGCTATTGAATTGATCAAGACAGCG	1087		
Qy	901	TTGACGCGCCATTCACCGCAAAAAACAGCGTATGTGTGACATTAACCACTTCAGTGCTG	960		
Db	1088	TTGACGCGCCATTCACCGCAAAAAACAGCGTATGTGTGACATTAACCACTTCAGTGCTG	1147		
Qy	961	TTTATGCGCGGACAGATATCTAATCTGGCAATCTCGCGCGCGACCTGAGAGCTCAACTGG	1020		
Db	1148	TTTATGCGCGGACAGATATCTAATCTGGCAATCTCGCGCGCGACCTGAGAGCTCAACTGG	1207		
Qy	1021	AAGCTTCCCGGTACACGGATTAACAACGCGCGACAGGTGGTGAATGCTGTGTAACGCTGG	1080		
Db	1208	AAGCTTCCCGGTACACGGATTAACAACGCGCGACAGGTGGTGAATGCTGTGTAACGCTGG	1267		
Qy	1081	CGTCCGCTAAGCGATTAACAGCCAGTGAATCAAGTTTCGTGCTTCCAGACTTTACAG	1140		
Db	1268	CGTCCGCTAAGCGATTAACAGCCAGTGAATCAAGTTTCGTGCTTCCAGACTTTACAG	1327		
Qy	1141	CAGATCGGTATTAACCGCGGTGTTCATTTAAATAGCCGCCCGGAGAGGTGAACCTAGCC	1200		
Db	1328	CAGATCGGTATTAACCGCGGTGTTCATTTAAATAGCCGCCCGGAGAGGTGAACCTAGCC	1387		
Qy	1201	CTGGCGAGATGTGAAGAGCGGAATGCGCAGGCACTGTGTTGCTGTGGCAGTTTACGCA	1260		
Db	1388	CTGGCGAGATGTGAAGAGCGGAATGCGCAGGCACTGTGTTGCTGTGGCAGTTTACGCA	1447		
Qy	1261	ATCGTGAATGAAGCAGCATACCCGCTTGCACTTTGTA	1299		
Db	1448	ATCGTGAATGAAGCAGCATACCCGCGGTGCACTTTGTA	1486		
RESULT 14	AF537219	1315 bp	DNA	linear	BCT 29-SEP-2002
LOCUS	AF537219				
DEFINITION	Escherichia coli AprrA (aprrA) gene, complete cds.				
ACCESSION	AF537219				
VERSION	AF537219.1	GI:23344056			
KEYWORDS	.				

SOURCE ORGANISM	Escherichia coli Escherichia coli Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE AUTHORS TITLE	1 (bases 1 to 1315) Zhang,Z., He,J., Yao,B., Zhou,Y., Chen,Y. and Yi,Y. Production of phytase and acid phosphatase by use of silkworm-bioreactor
JOURNAL REFERENCE AUTHORS TITLE	2 (bases 1 to 1315) Chen,Y., Zhu,Z., Zhang,Z. and He,J. Cloning and overexpression of phytase gene appa from Escherichia coli
JOURNAL REFERENCE AUTHORS TITLE	3 (bases 1 to 1315) Chen,Y., Zhu,Z., Zhang,Z. and He,J. Direct Submission Submitted (10-AUG-2002) Key Laboratory of Silkworm Biotechnology, Ministry of Agriculture, Sericultural Research Institute, Chinese Academy of Agricultural Sciences, Sibaidu, Zhenjiang, Jiangsu Province 212018, China
FEATURES	location/Qualifiers
source	1..1315 /organism="Escherichia coli" /mol_type="genomic DNA" /db_xref="taxon:562" 10..1308 /gene="appa" 10..1308 /note="phytase; phosphatase" /codon_start=1 /transl_table=1 /product="appa" /protein_id="AAN28334.1" /db_xref="GI:23344057" /translation="MKAILIPFLSLILPLFPOSAFQSBELKLESVIVISRGNRAAP TKAIQIMQDYPDAMPPTMMPVKLIGWLTFRGGELIYLIVHYGROPLVAGLAKKCPQSA QGVAILADVDRTRTKTGRBAFAGLAPCALIVHQADPSSPDLFNLKTVGCOLDNA NVTALISRAGGSIADPTGHRQIAFRELERYANPQSNLCIKRKPQESCTQALPS ELKTSADVNSLTGAVSLASMLTKLFLIQAGMPEPMGRITDSHQMNTLISLNAQF YLKSTPEVARSRAPIPLDLIKTALITPHIPPOKAYGVTLPTSVLFIAGHDTNLANIG ALEINMTLPQGPDMTPGEGLVFERMRLSDNSQMIQVSLVFQTLQWRDTPLSLNTW PPGRVKTLTLACGERRNAQMSLAFGTVINEARI PACSL"
gene	
CDS	
ORIGIN	
Query Match	99.4%; Score 1291; DB 1; Length 1315;
Best Local Similarity	99.6%; Pred. No. 0;
Matches 1294; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
QY	1 ATGAAAGCATTGATCCCATTTTATCTCTTGATTCGGTAAACCCGCAATTCGA 60       10 ATGAAAGCGATTTCATCCCATTTTATCTCTTGATTCGGTAAACCCGCAATTCGA 69
DB	61 TTGCGCTCAGAGTAGCGGAGCTGAAGCTGAAAGTGTGTGATTGTCACTGTCATGCT 120 70 TTGCGCTCAGAGTAGCGGAGCTGAAGCTGAAAGTGTGTGATTGTCACTGTCATGCT 129
QY	121 GTGCGTCTCCAAACCAAGGCCACGCACTGATGACGATGCCACCCGAGACGATGGCA 180 130 GTGCGTCTCCAAACCAAGGCCACGCACTGATGACGATGCCACCCGAGACGATGGCA 189
DB	181 AACTGGCGGTAAACCTGGGTTGGCTGACACCGCCGCGGTGTGACTAATCGCTATCTC 240       190 AACTGGCGGTAAACCTGGGTTGGCTGACACCGCCGCGGTGTGACTAATCGCTATCTC 249
QY	241 GGAATTACCAACGCCAGCGTCTGTAGCTGACGATGCTGCGGAAAAAGGCTGCCCCG 300 250 GGAATTACCAACGCCAGCGTCTGTAGCTGACGATGCTGCGGAAAAAGGCTGCCCCG 309
DB	301 CAGTCTGTCAGGTGCGCGATTATGCTGATGTCGACGACGTAACCGGTAAACAGGCGAA 360 310 CAGTCTGTCAGGTGCGCGATTATGCTGATGTCGACGACGTAACCGGTAAACAGGCGAA 369

QY 361 GCTTCGCGCGGCTGCGACCTGACCTGCAATTAACCGTACATACCCAGGACATACG 420  
 DB 370 GCTTCGCGCGGCTGCGACCTGACCTGCAATTAACCGTACATACCCAGGACATACG 429  
 QY 421 TCCAGTCCCGATCCGCTATTTAATCTCTAAAACCTGGGTTGGCAACGTGATACGGG 480  
 DB 430 TCCAGTCCCGATCCGCTATTTAATCTCTAAAACCTGGGTTGGCAACGTGATACGGG 489  
 QY 481 AACGTACTGACGCGATCTCTAGCAGGAGGAGAGGATCAATTGCTGACTTTACCGGCA 540  
 DB 490 AACGTACTGACGCGATCTCTAGCAGGAGGAGAGGATCAATTGCTGACTTTACCGGCA 549  
 QY 541 CGGCAAAACGCGCTTTGCGCAACCTGGAACCGGCTCTTAATTTTCCGCAATCAACTTGG 600  
 DB 550 CGGCAAAACGCGCTTTGCGCAACCTGGAACCGGCTCTTAATTTTCCGCAATCAACTTGG 609  
 QY 601 CTTAAACGTGAAGAAACAGAGCAAGCTGTCTAATTAACGAGGATTAACCATGGAATC 660  
 DB 610 CTTAAACGTGAAGAAACAGAGCAAGCTGTCTAATTAACGAGGATTAACCATGGAATC 669  
 QY 661 AAGTAGAGCGCGCAATGCTCTATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 720  
 DB 670 AAGTAGAGCGCGCAATGCTCTATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 729  
 QY 721 GAGATATTTCTCTGCAACAGAGCAAGGAAATCCGAGGCTGGGAGGAGATCAC 780  
 DB 730 AAGATATTTCTCTGCAACAGAGCAAGGAAATCCGAGGCTGGGAGGAGATCAC 789  
 QY 781 GATTCAACCAAGTGAACACCTTGTCTAAGTTTGCATTAACCGCAATTTATTTGTCACAA 840  
 DB 790 GATTCAACCAAGTGAACACCTTGTCTAAGTTTGCATTAACCGCAATTTATTTGTCACAA 849  
 QY 841 CGCAGCCAGAGGTTGGCCGCGCAGCCGCGCACCCTTATTAATTGATCAAGACACG 900  
 DB 850 CGCAGCCAGAGGTTGGCCGCGCAGCCGCGCACCCTTATTAATTGATCAAGACACG 909  
 QY 901 TTGAGCGCCCATCAACCGCAAAAACAGCGCTATGCTGACATTAACCATTCAGTCTG 960  
 DB 910 TTGAGCGCCCATCAACCGCAAAAACAGCGCTATGCTGACATTAACCATTCAGTCTG 969  
 QY 961 TTTATGCGCGGACACGATTAATCTGCGGCGGCGACAGGAGTCAATG 1020  
 DB 970 TTTATGCGCGGACACGATTAATCTGCGGCGGCGACAGGAGTCAATG 1029  
 QY 1021 ACCCTTCCCGGTACGCGGATTAACACGCGCAGAGTGTGAATGCTGTTGAACGCTGG 1080  
 DB 1030 ACCCTTCCCGGTACGCGGATTAACACGCGCAGAGTGTGAATGCTGTTGAACGCTGG 1089  
 QY 1081 CGTGGGTAAGCGATTAACAGCCAGTGAATCAAGTTTCGCTGTCTTCAAGACTTTACG 1140  
 DB 1090 CGTGGGTAAGCGATTAACAGCCAGTGAATCAAGTTTCGCTGTCTTCAAGACTTTACG 1149  
 QY 1141 CAGATGCGTGAATAAAGCGGCTGTCTAATTAATCGCGCGCGGAGAGGTAATGAC 1200  
 DB 1150 CAGATGCGTGAATAAAGCGGCTGTCTAATTAATCGCGCGCGGAGAGGTAATGAC 1209  
 QY 1201 CTGGCAGAGTGAAGAGCGAAATGCGCAGGAGCATGTTGTTGGAGGTTTACGCA 1260  
 DB 1210 CTGGCAGAGTGAAGAGCGAAATGCGCAGGAGCATGTTGTTGGAGGTTTACGCA 1269  
 QY 1261 ATCGTGAATGAAGACGATACCGGCTTGACGTTTGTAA 1299  
 DB 1270 ATCGTGAATGAAGACGATACCGGCTTGACGTTTGTAA 1308

RESULT 15  
 ECOLACPHC 1901 bp DNA linear BCT 04-Apr-2002  
 LOCUS Escherichia coli periplasmic phosphoanhydride phosphohydrolyase  
 DEFINITION (appa) gene, complete cds.  
 ACCESSION L03372  
 VERSION L03372.1 GI:145180

KEYWORDS acid phosphatase; periplasmic phosphoanhydride phosphohydrolyase.  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (sites)  
 AUTHORS Dassa, J., Marcq, C., and Boquet, P.L.  
 TITLE The complete nucleotide sequence of the Escherichia coli gene appA reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase  
 JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)  
 MEDLINE 90368616  
 PUBMED 2168385  
 REFERENCE 2 (bases 1 to 1901)  
 AUTHORS Ostalin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van Eften, R.L.  
 TITLE Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase  
 JOURNAL J. Biol. Chem. 267 (32), 22830-22836 (1992)  
 MEDLINE 93054596  
 PUBMED 1429631  
 FEATURES  
 source  
 1..1901  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 188..1486  
 /gene="appa"  
 188..1486  
 /gene="appa"  
 /standard\_name="acid phosphatase"  
 /citation\_1[1]  
 /codon\_start=1  
 /transl\_table=11  
 /product="periplasmic phosphoanhydride phosphohydrolyase"  
 /protein\_id="AA00004.1"  
 /db\_xref="GI:145181"  
 /translation="MKAILIIPLSILIPITLPQAFAPQSPBELKESVIVSRGVAP  
 TKATQMDQVTPDAMPPTWPKGLMLPRGGLIAYIGHYORQIVADGILAKKCPQS  
 GOVALIADVDERTKTEAFAGLAPDCAITVHTQDTSPPDLFPLRTGVQDLNA  
 NVYDALISRAAGSIADPTGRQTAFLERLVNFPQSNCLKREKQDESCSLTQALPS  
 ELKVSADNVSITGAVALSMLEIPLDQQAQWPEPMGRITDSHKQMDLSTLNAQF  
 YLLORTPEVARSRATPLDLIKTALPFPPOKQAYVTLPSTVLFLAGHDNLNANLGG  
 ALKINWTLPGQDNPPTPGGEIVPERWRRLSDNSQWIVFTLQQMDKPLSLNT  
 PEEVFLTAGCEBENAGKSLAGTQIVNEKRITACSL"  
 299..301  
 /gene="appa"  
 /citation\_1[1]  
 /phenotype="R20A mutant"  
 ORIGIN  
 Query Match 99.3%; Score 1290; DB 1; Length 1901;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 1293; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ATGAAAGCAATCTTAATCCATTTTATCTCTCTGATTCCTGTAACCCCGCAATCGCA 60  
 DB 188 ATGAAAGCAATCTTAATCCATTTTATCTCTCTGATTCCTGTAACCCCGCAATCGCA 247  
 QY 61 TTGGCTCAGAGTAGGCGGAGCTGAAGCTGGAAGTGTGATGTCAGTCTCATGT 120  
 DB 248 TTGGCTCAGAGTAGGCGGAGCTGAAGCTGGAAGTGTGATGTCAGTCTCATGT 307  
 QY 121 GTGCGTCTCAACCAAGGCAAGCACTGATGCAAGATGCAACCCGAGACGATGGCA 180  
 DB 308 GTGCGTCTCAACCAAGGCAAGCACTGATGCAAGATGCAACCCGAGACGATGGCA 367  
 QY 181 ACCTGGCCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGTGATGCTAATCGCTATCTC 240  
 DB 368 ACCTGGCCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGTGATGCTAATCGCTATCTC 427  
 QY 241 GGAATTAACCAAGCCGAGCTGTGTAGCGGACGATTTGCTGCGAAGAAAGGCGTGGCCG 300

```
Db 428 GGACATTACCAAGCCAGCGCTCTGTAGCCAGCGATTGCTGGCAAAAAGGCTGCCG 487
QY 301 CAGTCTGGTCAAGTTCGCGATTATTTGCTGATGTCAAGCGATACCGTTAAAAAGGGGAA 360
Db 488 CAGTCTGGTCAAGTTCGCGATTATTTGCTGATGTCAAGCGATACCGTTAAAAAGGGGAA 547
QY 361 GCCTTGCCGCGCGGCTGGCAGCTGATGCAATTAACGTACATACCCAGGAGATACG 420
Db 548 GCCTTGCCGCGCGGCTGGCAGCTGATGCAATTAACGTACATACCCAGGAGATACG 607
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTAAAAAACTGGCGTTTGCAACTGATTAACCG 480
Db 608 TCCAGTCCCGATCCGTTATTTATCTCTAAAAAACTGGCGTTTGCAACTGATTAACCG 667
QY 481 AACGTAGTGAACGCGATCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Db 668 AACGTAGTGAACGCGATCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 727
QY 541 CGGCAAAACGGCGGTTGGCGAAGTGAACGGGAGCTTAATTTTCCGCAATCAAACTTGTGC 600
Db 728 CGGCAAAACGGCGGTTGGCGAAGTGAACGGGAGCTTAATTTTCCGCAATCAAACTTGTGC 787
QY 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAAACGAGCAATTAACATCGAACTC 660
Db 788 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAAACGAGCAATTAACATCGAACTC 847
QY 661 AAGGTGAGCGCGGCAATGCTCATTTAACGGTGCAGTAAGCTCGCATCAATGCTGACG 720
Db 848 AAGGTGAGCGCGGCAATGCTCATTTAACGGTGCAGTAAGCTCGCATCAATGCTGACG 907
QY 721 GAGATATTTCTCTGCAACAAGCAAGGAAATGCCGAGCCGGGAGTGGGAAAGATCAC 780
Db 908 GAGATATTTCTCTGCAACAAGCAAGGAAATGCCGAGCCGGGAGTGGGAAAGATCAC 967
QY 781 GATTCAACCAAGTGAACACCTTGTGAATTTGATTAACGCGCAATTTTATTGCTACAA 840
Db 968 GATTCAACCAAGTGAACACCTTGTGAATTTGATTAACGCGCAATTTTATTGCTACAA 1027
QY 841 CGCAGCGCAGAGTTGCCCGAGCGCGCGCACCCCGTTATTGATTTGATCAAGACGCG 900
Db 1028 CGCAGCGCAGAGTTGCCCGAGCGCGCGCACCCCGTTATTGATTTGATCAAGACGCG 1087
QY 901 TTGAGCGCCCATCCACCGCAAAAACAGCGGATGATGATTAACCACTTCAGTCTG 960
Db 1088 TTGAGCGCCCATCCACCGCAAAAACAGCGGATGATGATTAACCACTTCAGTCTG 1147
QY 961 TTTATCGCGCGACAGATATCTGCAAAATCTCGCGGCGCACTGAGCTCACTGG 1020
Db 1148 TTTATCGCGCGACAGATATCTGCAAAATCTCGCGGCGCACTGAGCTCACTGG 1207
QY 1021 ACGCTTCCCGGATCAGCCGAGTAACAGCGCGCGAGTGTGAACGTGTTGAAAGCTGG 1080
Db 1208 ACGCTTCCCGGATCAGCCGAGTAACAGCGCGCGAGTGTGAACGTGTTGAAAGCTGG 1267
QY 1081 CGTCGGCTTAAGGATTAACAGCGAGTGAATTCAGTTCGCTGCTTCAGACTTTACG 1140
Db 1268 CGTCGGCTTAAGGATTAACAGCGAGTGAATTCAGTTCGCTGCTTCAGACTTTACG 1327
QY 1141 CAGATCGTGTATTAACCGCGCTGTATTAAATACCGCGCGGAGAGTGAACCTGACC 1200
Db 1328 CAGATCGTGTATTAACCGCGCTGTATTAAATACCGCGCGGAGAGTGAACCTGACC 1387
QY 1201 CTGGCAGAGTGTGAAGGAAATGCGCAGGCGATGTTCGTTGGCAGGTTTACGCA 1260
Db 1388 CTGGCAGAGTGTGAAGGAAATGCGCAGGCGATGTTCGTTGGCAGGTTTACGCA 1447
QY 1261 ATCGTGAATGAGACGATATCCGCTTGCAATTGTAA 1299
Db 1448 ATCGTGAATGAGACGATATCCGCTTGCAATTGTAA 1486
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 / Search time 885.787 Seconds  
(without alignments)  
8681.254 Million cell updates/sec

Title: US-09-926-375B-7\_COPY\_12653\_13951

Perfect score: 1299

Sequence: 1 atgaagcattcttaacc.....taccgcgttcgactgtgta 1299

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn19808:\*  
2: geneseqn19908:\*  
3: geneseqn20008:\*  
4: geneseqn20018:\*  
5: geneseqn20028:\*  
6: geneseqn20038:\*  
7: geneseqn20048:\*  
8: geneseqn20058:\*  
9: geneseqn20068:\*  
10: geneseqn20078:\*  
11: geneseqn20088:\*  
12: geneseqn20098:\*  
13: geneseqn20108:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	3470	3 AAC68298	AAC68298 R15/APPA
2	1299	100.0	4060	3 AAC68296	AAC68296 R15/APPA
3	1299	100.0	6116	3 AAC68297	AAC68297 R15/APPA
4	1299	100.0	6708	3 AAC68295	AAC68295 R15/APPA
5	1299	100.0	17732	3 AAC68300	AAC68300 Lama2/APP
6	1299	100.0	20623	3 AAC68294	AAC68294 Lama2/APP
7	1299	99.6	1299	3 ABK12514	ABK12514 DNA encod
8	1294.2	99.6	1299	3 ACA19297	ACA19297 Prokaryot
9	1294.2	99.6	1299	8 ADL16137	ADL16137 Escherich
10	1293.2	99.6	1901	6 AAD25463	AAD25463 Escherich
11	1293.2	99.6	1901	8 ADA19449	ADA19449 E. coli K
12	1293.2	99.6	1901	10 ADC87744	ADC87744 DNA encod
13	1293.2	99.6	1901	12 ADOS0297	ADOS0297 Escherich
14	1290	99.3	1901	12 ADOS0295	ADOS0295 Escherich
15	1288.4	99.2	1296	12 ADOS0296	ADOS0296 Escherich
16	1288	99.2	1296	12 ADL16139	ADL16139 Shigella
17	1287.8	99.1	1315	10 ADL02197	ADL02197 DNA encod
18	1286	98.7	5421	3 AAC68299	AAC68299 SV40/APPA
19	1282.2	98.7	1323	5 AAC88885	AAC88885 Escherich
20	1282.2	98.7	1323	6 AAD36473	AAD36473 Escherich

21	1282.2	98.7	1323	6 AAD25460	AAD25460 Escherich
22	1282.2	98.7	1323	8 ADA19445	ADA19445 E. coli B
23	1282.2	98.7	1323	12 ADOS0291	ADOS0291 Escherich
24	1282.2	98.7	1323	12 ADOS0301	ADOS0301 Kangaroo
25	1282.2	98.7	1323	12 ADOS0303	ADOS0303 Escherich
26	1268.8	97.7	1296	12 ADL16138	ADL16138 Shigella
27	1267.6	97.6	1901	8 ADA19452	ADA19452 E. coli K
28	1263.8	97.3	1486	4 AAD06831	AAD06831 E. coli a
29	1259.2	96.9	1108	10 ADC87742	ADC87742 DNA encod
30	1259.2	96.9	1108	12 ADOS0299	ADOS0299 Escherich
31	1257.4	96.8	1489	3 AAA28216	AAA28216 E. coli a
32	1257.4	96.8	1489	8 ACC57672	ACC57672 Escherich
33	1252.6	96.4	1486	4 AAD06832	AAD06832 E. coli a
34	1252.6	96.4	1486	4 ACC57673	ACC57673 Escherich
35	1170.2	90.1	1272	2 AAX26540	AAX26540 DNA encod
36	861.6	66.3	868	12 ADL16143	ADL16143 Proteus v
37	819.8	63.1	846	4 AAS41940	AAS41940 Genomic B
38	750.8	57.8	754	12 ADL16144	ADL16144 Zoo compo
39	749	57.7	757	12 ADL16142	ADL16142 Enterobac
40	744.2	57.3	757	12 ADL16141	ADL16141 Maize-opt
41	737.2	56.8	1256	9 AAD57147	AAD57147 Enterobac
42	735.2	56.6	1313	9 AAD57148	AAD57148 pNOV4054
43	734.2	56.5	1331	9 AAD57149	AAD57149 pNOV4058
44	722.4	55.6	724	12 ADL16140	ADL16140 Pasteurel
45	694.8	53.5	1737	5 AAS89885	AAS89885 DNA encod

## ALIGNMENTS

RESULT 1	AAC68298	standard; DNA; 3470 BP.
ID	AAC68298	
XX	AAC68298;	
AC	15-SRP-2003 (revised)	
DT	20-FEB-2001 (first entry)	
XX	R15/APPA plasmid coding sequence.	
DE	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;	
XX	environmental pollution; pig; ds.	
KW	Rattus sp.	
OS	Escherichia coli.	
OS	Chimeric.	
XX	WO200064247-A1.	
XX	02-NOV-2000.	
XX	20-APR-2000; 2000MO-CA000430.	
PF	23-APR-1999; 99US-0130508P.	
XX	(UYGU-) UNIV GUELPH.	
PA	Forsberg CW, Golovan S, Phillips JP;	
PI	WPI; 2000-687245/67.	
DR	P-PSDB; AAB36261.	
XX	Transgenic non-human animal for gastrointestinal tract specific	
PT	expression of a protein, preferably phytase, comprises a nucleic acid	
PT	sequence including a heterologous transgene construct encoding the	
PT	protein.	
XX	Claim 14; Fig 21; 152pp; English.	
PS	The present invention provides transgenic animals which produce desired	
CC	proteins, in this case pigs which express phytase in the salivary	
CC	gland. Low phytase production levels result in phytase in the diet being	



CC on 15-SEP-2003 to standardise OS field)

XX Sequence 4060 BP; 1257 A; 814 C; 843 G; 1146 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 1299; DB 3; Length 4060;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAGCATCTTAAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATCTGCA 60  
Db ATGAAAGCATCTTAAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATCTGCA 1870

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 120  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 1930

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 1871  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 1930

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 121  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 1931

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 181  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 1991

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 1991  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2050

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 241  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2051

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 301  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2111

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 361  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2171

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 421  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2231

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 481  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2291

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 541  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2351

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 601  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2411

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 661  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2471

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 721  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2531

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 781  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2591

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 841  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2651

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 901  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2711

Qy TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 961  
Db TTTCGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTCTCATGTGT 2770

Db 2771 TTATCGCCGGACACGATCTAATCTGCAAAATCTCGCGGCGACATGAGCTCAACTGG 2830

Qy 1021 ACGCTTCCCGGTGACGCGGATTAACACGCGCCGAGGTTGTAACCTGTTTGAACGCTGG 1080

Db 2831 ACGCTTCCCGGTGACGCGGATTAACACGCGCCGAGGTTGTAACCTGTTTGAACGCTGG 2890

Qy 1081 CPTCCGCTAAGGATTAACGCGGATTAACGCGGATTAACGCGGATTAACGCGGATTAACG 1140

Db 2891 CPTCCGCTAAGGATTAACGCGGATTAACGCGGATTAACGCGGATTAACGCGGATTAACG 2950

Qy 1141 CAGATCGTGAATTAACGCGGATTAACGCGGATTAACGCGGATTAACGCGGATTAACG 1200

Db 2951 CAGATCGTGAATTAACGCGGATTAACGCGGATTAACGCGGATTAACGCGGATTAACG 3010

Qy 1201 CTGGCAGATGTAAGAGGGAATGCGAGGAGGATGTTGTTGCGAGGATTTTACGCA 1260

Db 3011 CTGGCAGATGTAAGAGGGAATGCGAGGAGGATGTTGTTGCGAGGATTTTACGCA 3070

Qy 1261 ATGTGAATGAAGCAGCATACCCGCTTGCACTTTGTA 1299

Db 3071 ATGTGAATGAAGCAGCATACCCGCTTGCACTTTGTA 3109

RESULT 3  
AAC68297  
ID AAC68297 standard; DNA; 6116 BP.  
XX  
XX AAC68297;  
AC  
XX  
DT 15-SBP-2003 (revised)  
DT 20-FEB-2001 (first entry)  
XX  
XX R15/APPA plasmid coding sequence.  
DE  
XX  
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
KW environmental pollution; pig; ds.  
XX  
XX Rattus sp.  
OS Escherichia coli.  
OS Chimeric.  
XX  
PN NO200064247-A1.  
PD  
PD 02-NOV-2000.  
PD  
PF 20-APR-2000; 2000MO-CA000430.  
XX  
PR 23-APR-1999; 99US-0130508P.  
XX  
XX (UYGU-) UNIV GUELPH.  
XX  
XX Forsberg CW, Golovan S, Phillips JP;  
PI WPI; 2000-687245/67.  
XX  
XX P-PSDB; AAB36260.  
DR  
XX  
XX Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein.  
XX  
XX Claim 56; Fig 20; 152bp; English.  
XX  
XX The present invention provides transgenic animals which produce desired  
CC proteins, in this case pigs which express phytase in the salivary  
CC gland. Low phytase production levels result in phytate in the diet being  
CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 15-SEP-2003 to standardise OS field)

XX Sequence 6116 BP; 1724 A; 1386 C; 1407 G; 1599 T; 0 U; 0 Other;

Query Match 100.0%; Score 1299; DB 3; Length 6116;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAGCATCTTATCCATTTTATCTCTGATCCGTTAAACCCGCAATCTGCA 60  
 1811 ATGAAGCATCTTATCCATTTTATCTCTGATCCGTTAAACCCGCAATCTGCA 1870  
 QY 61 TTCGCTCAGAGTAGACCCGAGCTGAAGCTGAAAGTGTGATTTGTCAGTCATGCT 120  
 DB 1871 TTCGCTCAGAGTAGACCCGAGCTGAAGCTGAAAGTGTGATTTGTCAGTCATGCT 1930  
 QY 121 GTGCGTCTCCACCAAGCCGACGCACTGATGCAAGATGTACATCCCGACAGCGCA 180  
 DB 1931 GTGCGTCTCCACCAAGCCGACGCACTGATGCAAGATGTACATCCCGACAGCGCA 1990  
 QY 181 ACCTGCGCGGTAAACTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240  
 DB 1991 ACCTGCGCGGTAAACTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 2050  
 QY 241 GGACATTAACCAAGCCGAGCTGTGATGCGCAAGATTGCTGGCAAAAAGGCTGCCG 300  
 DB 2051 GGACATTAACCAAGCCGAGCTGTGATGCGCAAGATTGCTGGCAAAAAGGCTGCCG 2110  
 QY 301 CAGTCTGTCAAGTGCAGCTTATTTGCTGATGTGACAGACCTTACCCTGTAAACAGGCCGA 360  
 DB 2111 CAGTCTGTCAAGTGCAGCTTATTTGCTGATGTGACAGACCTTACCCTGTAAACAGGCCGA 2170  
 QY 361 GCGTTCGCGCGCGGGCTGGCACTGACTGTGCAATACGCTACATTCACGAGCAGATAG 420  
 DB 2171 GCGTTCGCGCGCGGGCTGGCACTGACTGTGCAATACGCTACATTCACGAGCAGATAG 2230  
 QY 421 TCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCGTTGCCAATGATTAACGCG 480  
 DB 2231 TCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCGTTGCCAATGATTAACGCG 2290  
 QY 481 AACGTACTGACGCGCATCTCTCAGCAGGAGGAGGCTCAATTGCTGACCTTTACCGCGGCAT 540  
 DB 2291 AACGTACTGACGCGCATCTCTCAGCAGGAGGAGGCTCAATTGCTGACCTTTACCGCGGCAT 2350  
 QY 541 CGGCAAAACGCGCTTTTGGCGAACTGAAACGAGGCTCTTAATTTTCCGCAATCAAACTTGTGC 600  
 DB 2351 CGGCAAAACGCGCTTTTGGCGAACTGAAACGAGGCTCTTAATTTTCCGCAATCAAACTTGTGC 2410  
 QY 601 CTTAAACGTGAAACAGGACGAAAGCTGTCTTAACGCAAGCATTTACCATCGAATCTC 660  
 DB 2411 CTTAAACGTGAAACAGGACGAAAGCTGTCTTAACGCAAGCATTTACCATCGAATCTC 2470  
 QY 661 AAGGTAGGCGCGCAATGTCTCATTTAACCGGTGCGGTAAAGCTCGCATCAATGCTGAG 720  
 DB 2471 AAGGTAGGCGCGCAATGTCTCATTTAACCGGTGCGGTAAAGCTCGCATCAATGCTGAG 2530  
 QY 721 GAGATATTTCTCTGCAACAGACAGGAAATCCGAGCGCGGAGTGGGAAAGATCACC 780  
 DB 2531 GAGATATTTCTCTGCAACAGACAGGAAATCCGAGCGCGGAGTGGGAAAGATCACC 2590  
 QY 781 GATTCAACACAGTGAACACCTTGTAGTTTGCAATACGCGCAATTTTATTTGCTACAA 840  
 DB 2591 GATTCAACACAGTGAACACCTTGTAGTTTGCAATACGCGCAATTTTATTTGCTACAA 2650  
 QY 841 CGACGCGCAGAGTTCGCGCAGCCGCGCAACCCCGTTATTTAGATTTGATCAAGACAGG 900  
 DB 2651 CGACGCGCAGAGTTCGCGCAGCCGCGCAACCCCGTTATTTAGATTTGATCAAGACAGG 2710  
 QY 901 TTGACGCGCCATCCACGCAAAAACAGGCGTATGTGTGATTAATCCACTTCAAGTCTG 960  
 DB 2711 TTGACGCGCCATCCACGCAAAAACAGGCGTATGTGTGATTAATCCACTTCAAGTCTG 2770  
 QY 961 TTTATGCGCGGACAGATTAATCTGGCAATCTGGCGCGCACTGAGCTCAACTGG 1020  
 DB 2771 TTTATGCGCGGACAGATTAATCTGGCAATCTGGCGCGCACTGAGCTCAACTGG 2830

QY 1021 ACGCTTCCCGGTCAAGCCGATTAACACGCGCCAGGTGTGAACTGTGTTGAACGCTGG 1080  
 DB 2831 ACGCTTCCCGGTCAAGCCGATTAACACGCGCCAGGTGTGAACTGTGTTGAACGCTGG 2890  
 QY 1081 CGTGGCTTAAGCGATTAACGACGAGTGTGATTAAGTTTGGCTGTCTTCAACATTTACG 1140  
 DB 2891 CGTGGCTTAAGCGATTAACGACGAGTGTGATTAAGTTTGGCTGTCTTCAACATTTACG 2950  
 QY 1141 CAGATGCGGTAAACGCGCGCTGTCAATTAATACGCGCCGAGAGAGTGAACCTGACC 1200  
 DB 2951 CAGATGCGGTAAACGCGCGCTGTCAATTAATACGCGCCGAGAGAGTGAACCTGACC 3010  
 QY 1201 CTGGCAGATGTGAAGACGAATGCGAGGCAATGTGTTGTTGGCAGGTTTACGCA 1260  
 DB 3011 CTGGCAGATGTGAAGACGAATGCGAGGCAATGTGTTGTTGGCAGGTTTACGCA 3070  
 QY 1261 ATCGTGAATGAAGACGCAATACCGCTTGCAGTTTGTAA 1299  
 DB 3071 ATCGTGAATGAAGACGCAATACCGCTTGCAGTTTGTAA 3109

RESULT 4  
 AAC68295  
 ID AAC68295 standard; DNA; 6708 BP.  
 XX  
 AC AAC68295;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 20-FEB-2001 (first entry)  
 XX  
 DE R15/APPA plasmid coding sequence.  
 XX  
 KM Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KM environmental pollution; pig; ds.  
 XX  
 OS Rattus sp.  
 OS Escherichia coli.  
 OS Chimeric.  
 PN WO20064247-A1.  
 XX  
 XX 02-NOV-2000.  
 PD  
 XX 20-APR-2000; 2006WO-CA000430.  
 PF  
 XX 23-APR-1999; 99US-0130508P.  
 ER  
 PA (UYGU-) UNIV GUELPH.  
 XX  
 PI Forsberg CW, Golovan S, Phillips JP;  
 DR WPI; 2000-687245/67.  
 XX  
 DR P-PSDB; AAB36258.  
 XX  
 PT Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprising a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.  
 XX  
 PS Claim 56; Fig 18; 1522P; English.  
 XX  
 CC The present invention provides transgenic animals which produce desired  
 CC proteins, in this case pigs which expresses phytase in the salivary  
 CC gland. Low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
 CC on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 6708 BP; 1916 A; 1479 C; 1515 G; 1798 T; 0 U; 0 Other;

Query Match 100.0%; Score 1299; DB 3; Length 6708;  
 Best Local Similarity 100.0%; Pred. No. 0;



Matches	1299;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	ATGAAAGCCATCTTAATCCCATTTTATCTTTCTGATTCGGTAAACCCCGCAATCTGCA	60						
Db	1811	ATGAAAGCCATCTTAATCCCATTTTATCTTTCTGATTCGGTAAACCCCGCAATCTGCA	1870						
Qy	61	TTTCGCTCAAGTGAAGCCGAGCTGAAAGTGTGTGATTTTCTAGTCCGTCATGTGT	120						
Db	1871	TTTCGCTCAAGTGAAGCCGAGCTGAAAGTGTGTGATTTTCTAGTCCGTCATGTGT	1930						
Qy	121	GTGCGGTCTCAACCAAGGCCACGCAACTGATGCAAGATGTCACCCGAGCGATGGCA	180						
Db	1931	GTGCGGTCTCAACCAAGGCCACGCAACTGATGCAAGATGTCACCCGAGCGATGGCA	1990						
Qy	181	ACCTGGCCGGTAAACCTGGGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC	240						
Db	1991	ACCTGGCCGGTAAACCTGGGTGTGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC	2050						
Qy	241	GGACATTAACCAAGCCGAGCTGTGTAGCCGAGATTTGCGCGCAAAAAGGGCTGCGC	300						
Db	2051	GGACATTAACCAAGCCGAGCTGTGTAGCCGAGATTTGCGCGCAAAAAGGGCTGCGC	2110						
Qy	301	CAGCTGTGAGGTTCGGATTAATGCTGATGTCGACGAGCGGTACCCGTAACAGCGGAA	360						
Db	2111	CAGCTGTGAGGTTCGGATTAATGCTGATGTCGACGAGCGGTACCCGTAACAGCGGAA	2170						
Qy	361	GCCTTCGCGCGCGGGCTGCGACCTGACTGTGCAATAACCTGTACATACCAAGGATACG	420						
Db	2171	GCCTTCGCGCGCGGGCTGCGACCTGACTGTGCAATAACCTGTACATACCAAGGATACG	2230						
Qy	421	TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCTTGCCCACTGGATTAACGCG	480						
Db	2231	TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGCTTGCCCACTGGATTAACGCG	2290						
Qy	481	AAGGTGACGACGCGATCTCAGACAGGGCAGAGGGCTCAATTGCTGACTTTACCGGGCAT	540						
Db	2291	AAGGTGACGACGCGATCTCAGACAGGGCAGAGGGCTCAATTGCTGACTTTACCGGGCAT	2350						
Qy	541	CGGCAACCGCGCTTTCCGCAACTGGAAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC	600						
Db	2351	CGGCAACCGCGCTTTCCGCAACTGGAAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC	2410						
Qy	601	CTTAAACGTGAAACGAGCAAGAACTGTTCAATTAACGAGGATTAACCATCGGAATCTC	660						
Db	2411	CTTAAACGTGAAACGAGCAAGAACTGTTCAATTAACGAGGATTAACCATCGGAATCTC	2470						
Qy	661	AAGGTGAGCGCGCAATGTTCTCATTAACGGGTGCGGTAAGCTCGCATCAATGCTGACG	720						
Db	2471	AAGGTGAGCGCGCAATGTTCTCATTAACGGGTGCGGTAAGCTCGCATCAATGCTGACG	2530						
Qy	721	GAGATATTTCTCTGCAACAGCAACAGGAATGCGGAGCGCGGGTGGGGAAGGATACCC	780						
Db	2531	GAGATATTTCTCTGCAACAGCAACAGGAATGCGGAGCGCGGGTGGGGAAGGATACCC	2590						
Qy	781	GATTACACCAAGTGGAAACACTTGTCTAAGTTTGATTAACGCGCAATTTTATTTGCTACAA	840						
Db	2591	GATTACACCAAGTGGAAACACTTGTCTAAGTTTGATTAACGCGCAATTTTATTTGCTACAA	2650						
Qy	841	CGACGCGCAGAGGTGGCCGAGCGCGGCAACCCGTTATTAAGTTTATGATCAAGACGCG	900						
Db	2651	CGACGCGCAGAGGTGGCCGAGCGCGGCAACCCGTTATTAAGTTTATGATCAAGACGCG	2710						
Qy	901	TTGACGCGCCATCCACCGCAAAAACAGGCGGTATGTGACATTAACCATCTTCAGTGTG	960						
Db	2711	TTGACGCGCCATCCACCGCAAAAACAGGCGGTATGTGACATTAACCATCTTCAGTGTG	2770						
Qy	961	TTTATGCGCGGACACGATCTAATCTGCAAAATCTCGCGCGCGCATGAGACTCAACTGCG	1020						
Db	2771	TTTATGCGCGGACACGATCTAATCTGCAAAATCTCGCGCGCGCATGAGACTCAACTGCG	2830						
Qy	1021	ACGCTTCGCGGTACGCGGATTAACGCGCGCAGGTGTGTAACGTGTGTTGAACGCTGCG	1080						
Db	2831	ACGCTTCGCGGTACGCGGATTAACGCGCGCAGGTGTGTAACGTGTGTTGAACGCTGCG	2890						

Qy	1081	CGTCGGCTAAGCGGTAACAGCCAGTGGATTGAGTTTGGCTGTCTTCAGACTTTACAG	1140
Db	2891	CGTCGGCTAAGCGGTAACAGCCAGTGGATTGAGTTTGGCTGTCTTCAGACTTTACAG	2950
Qy	1141	CAGATGCGGTATTAACCGCGCTGTCTTCAATTAATACGCGCGCGAGAGGTGAACCTGACC	1200
Db	2951	CAGATGCGGTATTAACCGCGCTGTCTTCAATTAATACGCGCGCGAGAGGTGAACCTGACC	3010
Qy	1201	CTGGCAGATGTGAAGCGGAATGCGCAGGCGCATGTGTTCTGTTGGAGGTTTTTACGCCAA	1260
Db	3011	CTGGCAGATGTGAAGCGGAATGCGCAGGCGCATGTGTTCTGTTGGAGGTTTTTACGCCAA	3070
Qy	1261	ATCGTAATGAAGCAGCATACCCGCTTGCAAGTTGTAA	1299
Db	3071	ATCGTAATGAAGCAGCATACCCGCTTGCAAGTTGTAA	3109
RESULT 5			
AAC68300			
ID	AAC68300 standard; DNA; 17732 BP.		
AC	AAC68300;		
XX			
DT	15-SBP-2003 (revised)		
DT	20-FEB-2001 (first entry)		
XX			
DE	Lama2/APPA plasmid coding sequence.		
XX			
KM	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;		
KM	environmental pollution; pig; ds.		
OS	Mus musculus.		
OS	Escherichia coli.		
OS	Chimeric.		
PN	W0200064247-A1.		
XX			
PD	02-NOV-2000.		
XX			
PF	20-APR-2000; 2000WO-CA000430.		
XX			
PR	23-APR-1999; 99US-0130508P.		
XX			
PA	(UYGU-) UNIV GUEBLPH.		
XX			
PI	Forsberg CW, Golovan S, Phillips JP;		
XX			
DR	WPI; 2000-687245/67.		
XX			
PT	Transgenic non-human animal for gastrointestinal tract specific		
PT	expression of a protein, preferably phytase, comprises a nucleic acid		
PT	sequence including a heterologous transgene construct encoding the		
PT	protein.		
XX			
PS	Claim 14; Fig 23; 152pp; English.		
XX			
CC	The present invention provides transgenic animals which produce desired		
CC	proteins, in this case pigs which expresses phytase in the salivary		
CC	gland. Low phytase production levels result in phytate in the diet being		
CC	excreted and causing phosphorus contamination in water, as well as		
CC	reducing the growth of animals. The invention provides a number of		
CC	transgenes containing the E. coli APPA phytase coding sequence. (Updated		
CC	on 15-SBP-2003 to standardise OS field)		
XX			
SO	Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;		
Query Match 100.0%; Score 1299; DB 3; Length 17732;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ATGAAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTAAACCCCGCAATCTGCA	60

```

Db      12653 ATGAAAGCATCTTAATCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATCTGCA 12712
Qy      61 TTGCTCAGAGTAGAGCCGAGCTGAAAGCTGAAAGTGTGTGATTTGTCAGTGTGATGCT 120
Db      12713 TTGCTCAGAGTAGAGCCGAGCTGAAAGCTGAAAGTGTGTGATTTGTCAGTGTGATGCT 12772
Qy      121 GTGCGTCTCCAAACCAAGGCCCACTGATGCAAGATGTCAACCCCAAGCCATGGCCA 180
Db      12773 GTGCGTCTCCAAACCAAGGCCCACTGATGCAAGATGTCAACCCCAAGCCATGGCCA 12832
Qy      181 ACCTGGCCGGTAAACTGGGTGTGCTGACACCGCGGTGTGAGCTATCTC 240
Db      12833 ACCTGGCCGGTAAACTGGGTGTGCTGACACCGCGGTGTGAGCTATCTC 12892
Qy      241 GGACATTAACCAAGCCAGCGTGTGTGAGCCGACGATTCGTGCGCAAAAAGGCTGCCCC 300
Db      12893 GGACATTAACCAAGCCAGCGTGTGTGAGCCGACGATTCGTGCGCAAAAAGGCTGCCCC 12952
Qy      301 CAGTCTGCTCAGGTGCGCATTTATGCTGATGTGACGAGCGTAAACCGTAAACAGGCGCA 360
Db      12953 CAGTCTGCTCAGGTGCGCATTTATGCTGATGTGACGAGCGTAAACCGTAAACAGGCGCA 13012
Qy      361 GCCTTGCCGCGGCGGTGCACTGCTGCAATTAACCGTACATTAACCGAGCAGATACG 420
Db      13013 GCCTTGCCGCGGCGGTGCACTGCTGCAATTAACCGTACATTAACCGAGCAGATACG 13072
Qy      421 TCCAGTCCCGATCCGTTATTTAATCCTCTAATAAACTGGGCTTGCCAACTGATTAACGG 480
Db      13073 TCCAGTCCCGATCCGTTATTTAATCCTCTAATAAACTGGGCTTGCCAACTGATTAACGG 13132
Qy      481 AACGTGATCGAGCGCATCTCTGACGAGGAGGAGGCTCAATTGCTGACTTTACCGGCGCAT 540
Db      13133 AACGTGATCGAGCGCATCTCTGACGAGGAGGAGGCTCAATTGCTGACTTTACCGGCGCAT 13192
Qy      541 CGGCAAAACGGCGTTTTCGCAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC 600
Db      13193 CGGCAAAACGGCGTTTTCGCAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC 13252
Qy      601 CTTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTACATGGAACATC 660
Db      13253 CTTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTACATGGAACATC 13312
Qy      661 AAGGTAGGCGCGCAATGCTCATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGAGC 720
Db      13313 AAGGTAGGCGCGCAATGCTCATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGAGC 13372
Qy      721 GAGATATTTCTCTGCAACAGACAGGAAATGCGAGCCGAGTGTGGGAAAGATCAC 780
Db      13373 GAGATATTTCTCTGCAACAGACAGGAAATGCGAGCCGAGTGTGGGAAAGATCAC 13432
Qy      781 GATTACACAGAGTGAACACCTTGTCTAAGTTGTGATTAACGCGCAATTTTATTTGCTACAA 840
Db      13433 GATTACACAGAGTGAACACCTTGTCTAAGTTGTGATTAACGCGCAATTTTATTTGCTACAA 13492
Qy      841 CGACGCGCAGAGGTGGCCGAGCCGCGCAACCCGTTATTAATTTGATCAAGACAGC 900
Db      13493 CGACGCGCAGAGGTGGCCGAGCCGCGCAACCCGTTATTAATTTGATCAAGACAGC 13552
Qy      901 TTGACGCGCCATCCACGCAAAAACAGCGCTATGTGTGACATTAACCATTCAGTGTG 960
Db      13553 TTGACGCGCCATCCACGCAAAAACAGCGCTATGTGTGACATTAACCATTCAGTGTG 13612
Qy      961 TTATATGCGCGGACACGATCTAATCTGGCAAAATCTGGCGCGGCACTGAGGCTCACTGG 1020
Db      13613 TTATATGCGCGGACACGATCTAATCTGGCAAAATCTGGCGCGGCACTGAGGCTCACTGG 13672
Qy      1021 ACCGTTCCGGTCAAGCGGATTAACAGCGCGCAAGTGTGAACCTGGTTGAACGCTGG 1080
Db      13673 ACCGTTCCGGTCAAGCGGATTAACAGCGCGCAAGTGTGAACCTGGTTGAACGCTGG 13732
Qy      1081 CGTCGCTAAGCATTAACAGCCAGTGTGATTCAGGTTTCCGTGCTTCAGACTTTACAG 1140

```

```

Db      13733 CGTCGCTAAGCATTAACAGCCAGTGAATTCAGGTTTCGTGTGCTTCCAGACTTTACG 13792
Qy      1141 CAGATGCTGTATTAACGCGCTGTCAATTAATACGCGCCCGAGAGGTGAACCTGACC 1200
Db      13793 CAGATGCTGTATTAACGCGCTGTCAATTAATACGCGCCCGAGAGGTGAACCTGACC 13852
Qy      1201 CTGCGCAGATGTGAAGACGCAAAATGCGCAGGCGCATGTGTGTTGGCAGGTTTACGCA 1260
Db      13853 CTGCGCAGATGTGAAGACGCAAAATGCGCAGGCGCATGTGTGTTGGCAGGTTTACGCA 13912
Qy      1261 ATCGTAATGAAGCAGCATACCCGCTTGCAATTGTTGA 1299
Db      13913 ATCGTAATGAAGCAGCATACCCGCTTGCAATTGTTGA 13951

```

RESULT 6  
AAC68294  
ID AAC68294 standard; DNA; 20623 BP.

XX AAC68294;

AC 15-SEP-2003 (revised)

DT 20-FEB-2001 (first entry)

DE Lama2/APPA plasmid coding sequence.

KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth; environmental pollution; pig; ds.

OS Mus musculus.

OS Escherichia coli.

XX Chimeric.

PN WO200064247-A1.

XX 02-NOV-2000.

PD 20-APR-2000; 2000WO-CA000430.

PF 23-APR-1999; 99US-0130508P.

PA (UYGU-) UNIV GUELPH.

PI Forsberg CW, Golovan S, Phillips JP;

XX WPI; 2000-687245/67.

DR P-PSDB; AAB36257.

PT Transgenic non-human animal for gastrointestinal tract specific

XX expression of a protein, preferably phytase, comprises a nucleic acid

PT sequence including a heterologous transgene construct encoding the

XX protein.

PS Claim 56; Fig 5; 152pp; English.

XX The present invention provides transgenic animals which produce desired

CC proteins, in this case pigs which expresses phytase in the salivary

CC gland. Low phytase production levels result in phytate in the diet being

CC excreted and causing phosphorus contamination in water, as well as

CC reducing the growth of animals. The invention provides a number of

CC transgenes containing the E. coli APPA phytase coding sequence. (Updated

CC on 15-SEP-2003 to standardise OS field)

XX Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

SO Query Match 100.0%; Score 1299; DB 3; Length 20623;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAGCATCTTAATCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATCTGCA 60

Db 12653 ATGAAAGCATCTTAATCCATTTTATCTCTTCGATTCCTGTAACCCCGCAATCTGCA 12712



Db 181 ACCCTGCGCGTAAACTGAGTGGTGGCTGACACCGCGGCTGGTGAAGTAAATGCTATCTC 240  
QY 241 GGACATTAACCAAGCCAGCGTCTGGTAGCCGAGATTGCTGGCAAAAAAGGCTCCCG 300  
Db 241 GGACATTAACCAAGCCAGCGTCTGGTAGCCGAGATTGCTGGCAAAAAAGGCTCCCG 300  
QY 301 CAGTCTGGTCAAGTGGCGATTTATGCTGATGTCGACGAGCCGTAACCCGTAACAGGCGGA 360  
Db 301 CAGTCTGGTCAAGTGGCGATTTATGCTGATGTCGACGAGCCGTAACCCGTAACAGGCGGA 360  
QY 361 GCTTCGCGCGGCGCTGACCTGACTGTGAATTAACCGTACATACCAGGCAATACG 420  
Db 361 GCTTCGCGCGGCGCTGACCTGACTGTGAATTAACCGTACATACCAGGCAATACG 420  
QY 421 TCCAGTCCCGGATCCGTTATTTAATCTCTAATAAACTGCGCTTCCCACTGATTAACGCG 480  
Db 421 TCCAGTCCCGGATCCGTTATTTAATCTCTAATAAACTGCGCTTCCCACTGATTAACGCG 480  
QY 481 AACGTGACTGACGCGATCTCTACAGAGGCAAGAGGTCAATTGCTGACTTTACCGGCGAT 540  
Db 481 AACGTGACTGACGCGATCTCTACAGAGGCAAGAGGTCAATTGCTGACTTTACCGGCGAT 540  
QY 541 CGGCAAAACGCGCTTGGCCGAACTGGAAAGGAGTCTTAATTTTCCGCAATCAAACTTGTGC 600  
Db 541 CGGCAAAACGCGCTTGGCCGAACTGGAAAGGAGTCTTAATTTTCCGCAATCAAACTTGTGC 600  
QY 601 CTTAAACGTGAGAAACAGAGCAGAACTGTCTTAATTAACGACGATTAACCATCGAATC 660  
Db 601 CTTAAACGTGAGAAACAGAGCAGAACTGTCTTAATTAACGACGATTAACCATCGAATC 660  
QY 661 AAGGTGAGGCGGCAATGCTCTAATTAACGCGGTGTAAGCTGCAATCAATGCTGAGG 720  
Db 661 AAGGTGAGGCGGCAATGCTCTAATTAACGCGGTGTAAGCTGCAATCAATGCTGAGG 720  
QY 721 GAGATATTTCTCTGCAACAGACAGGAAATGCGGACCGGCGTGGGAAAGATCAC 780  
Db 721 GAGATATTTCTCTGCAACAGACAGGAAATGCGGACCGGCGTGGGAAAGATCAC 780  
QY 781 GATTCAACACGATGGAACACTTGTCTAATGTTGCAATTAAGCGCAATTTATTTGCTACAA 840  
Db 781 GATTCAACACGATGGAACACTTGTCTAATGTTGCAATTAAGCGCAATTTATTTGCTACAA 840  
QY 841 CGACGCGCAGAGGTGCGCGACCGCGCCACCCCGTTATTAATTAATTAATTAATTAATTA 900  
Db 841 CGACGCGCAGAGGTGCGCGACCGCGCCACCCCGTTATTAATTAATTAATTAATTAATTA 900  
QY 901 TTGACGCCCATCAACCGCAAAACAGGCGTATGCTGATGATTAATTAATTAATTAATTA 960  
Db 901 TTGACGCCCATCAACCGCAAAACAGGCGTATGCTGATGATTAATTAATTAATTAATTA 960  
QY 961 TTTATGCGCGGACAGATTAATTAATGCGCAAAATCTGCGCGCGCACTGAGCTCAACTG 1020  
Db 961 TTTATGCGCGGACAGATTAATTAATGCGCAAAATCTGCGCGCGCACTGAGCTCAACTG 1020  
QY 1021 AGCGTCCCGGTGACCGGATTAACGCGCGCGAGTGTGAATGCGGTGTTGAACCTG 1080  
Db 1021 AGCGTCCCGGTGACCGGATTAACGCGCGCGAGTGTGAATGCGGTGTTGAACCTG 1080  
QY 1081 CGTGGGTAAAGCATTAACAGCCAGTGAATCAAGTTTCGCTGCTTCCAGACTTTACAG 1140  
Db 1081 CGTGGGTAAAGCATTAACAGCCAGTGAATCAAGTTTCGCTGCTTCCAGACTTTACAG 1140  
QY 1141 CAGATGCGGTAAATTAACGCGCTGCTCAATTAATACGCGCGCGGAGAGGTGAATGAC 1200  
Db 1141 CAGATGCGGTAAATTAACGCGCTGCTCAATTAATACGCGCGCGGAGAGGTGAATGAC 1200  
QY 1201 CTGGCAGATGTAAGAGCAAAATGCGCAGGAGATGTTGCTTGGAGAGTTTACGCA 1260  
Db 1201 CTGGCAGATGTAAGAGCAAAATGCGCAGGAGATGTTGCTTGGAGAGTTTACGCA 1260  
QY 1261 ATCGTAATGAGACGCAATACCGCTTGGAGTTTGA 1299  
Db 1261 ATCGTAATGAGACGCAATACCGCTTGGAGTTTGA 1299

Db 1261 ATCGTAATGAGACGCAATACCGCGCTGACGTTTGA 1299  
RESULT 8  
ACAI9297  
ID ACAI9297 standard; DNA; 1299 BP.  
XX  
XX ACAI9297;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #954.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
XX Escherichia coli.  
XX  
XX WO20027183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX  
XX 06-SEP-2001; 2001US-00948993.  
XX  
XX 25-OCT-2001; 2001US-03492923P.  
XX  
XX 08-FEB-2002; 2002US-00072851.  
XX  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX MPI; 2003-029926/02.  
XX  
XX P-PSDB; ABU15427.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 7167; 1766bp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation; or the activity of a gene in an operon required for  
XX cell proliferation; (7) identifying a compound that influences the activity  
XX of the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than S. aureus, S. typhimurium,  
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
XX prokaryotic essential genes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in

CC electronic format directly from WIFO at  
CC ftp.wifo.int/pub/published\_pcr\_sequences  
XX

Sequence 1299 BP; 317 A; 345 C; 355 G; 282 T; 0 U; 0 Other;

Query Match 99.6%; Score 1294.2; DB 8; Length 1299;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATGAAAGCATCTTAATCCCATTTTATCTCTGATTCCTGTAACCCCGCAATCTGCA 60
DB 1 ATGAAAGCATCTTAATCCCATTTTATCTCTGATTCCTGTAACCCCGCAATCTGCA 60
QY 61 TTGCTCAGAGTGAAGCCGAGCTGAAAGTGTGTGATTTGTGATGCTGATGCT 120
DB 61 TTGCTCAGAGTGAAGCCGAGCTGAAAGTGTGTGATTTGTGATGCTGATGCT 120
QY 121 GTGCGGTCTCAACCAAGGCGAAGCTGATGCAAGATGTCACCCGAGCATGAGCA 180
DB 121 GTGCGGTCTCAACCAAGGCGAAGCTGATGCAAGATGTCACCCGAGCATGAGCA 180
QY 181 ACCTGCGCGGTAAATCTGGGTTGGCTGACACCGCGGTTGTGAGCTAATCGCTATCTC 240
DB 181 ACCTGCGCGGTAAATCTGGGTTGGCTGACACCGCGGTTGTGAGCTAATCGCTATCTC 240
QY 241 GGAATTACCAACGCGAGCTGTGTGACCGAGATTGCTGCGGAAAAAGGCTGCCG 300
DB 241 GGAATTACCAACGCGAGCTGTGTGACCGAGATTGCTGCGGAAAAAGGCTGCCG 300
QY 301 CAGTCTGTGTCAGGTCCGATTAATTTGCTGATGTCAGAGGTACCCGTAACAGAGGGA 360
DB 301 CAGTCTGTGTCAGGTCCGATTAATTTGCTGATGTCAGAGGTACCCGTAACAGAGGGA 360
QY 361 GCCTTGCGCGCGGGCTGACCTGACTGTGCAATAACCTGATACCCAGGAGATACG 420
DB 361 GCCTTGCGCGCGGGCTGACCTGACTGTGCAATAACCTGATACCCAGGAGATACG 420
QY 421 TCCAGTCCCGATCCGTTATTTAAATCTCTTAAAACTGGCTTCCCACTGATTAACGCG 480
DB 421 TCCAGTCCCGATCCGTTATTTAAATCTCTTAAAACTGGCTTCCCACTGATTAACGCG 480
QY 481 AAGTGAATGACCGGATCTCTCAGACGGGCGAGGCTCATTTGCTGATTCAGGCGAT 540
DB 481 AAGTGAATGACCGGATCTCTCAGACGGGCGAGGCTCATTTGCTGATTCAGGCGAT 540
QY 541 CGGCAAAACGCGCTTTGCGCAACTGGAACGGGCTGTTAATTTCCGCAATCAACTTGTGC 600
DB 541 CGGCAAAACGCGCTTTGCGCAACTGGAACGGGCTGTTAATTTCCGCAATCAACTTGTGC 600
QY 601 CTTAAACGTGAGAAACAGACGAAGCTGTTCAATTAACGAGGATTAACATCGGAATCTC 660
DB 601 CTTAAACGTGAGAAACAGACGAAGCTGTTCAATTAACGAGGATTAACATCGGAATCTC 660
QY 661 AAGGTGCGCGCGAATGTTCTCTTAACCGGTGCGGTAAGCTCCGATCAAGTGTAGC 720
DB 661 AAGGTGCGCGCGAATGTTCTCTTAACCGGTGCGGTAAGCTCCGATCAAGTGTAGC 720
QY 721 GAGATATTTTCTCTGCAACAGACAGAGGATGCGGAGCGGAGTGGAGGATCAAC 780
DB 721 GAGATATTTTCTCTGCAACAGACAGAGGATGCGGAGCGGAGTGGAGGATCAAC 780
QY 781 GATTACACCAAGTGAACACCTTCTGTAAGTTGCAATTAACGCGCAATTTTATTTGCTCA 840
DB 781 GATTACACCAAGTGAACACCTTCTGTAAGTTGCAATTAACGCGCAATTTTATTTGCTCA 840
QY 841 CGCAGCGCAAGTGTGCGCGAGCGCGCGCAACCCGTTATTAAGTTGAATCAACAGCG 900
DB 841 CGCAGCGCAAGTGTGCGCGAGCGCGCGCAACCCGTTATTAAGTTGAATCAACAGCG 900
QY 901 TTGACGCGCGCATCAACCGCAAAAACAGGCGATGATGTGATTAACCATTCAGTGTCTG 960
DB 901 TTGACGCGCGCATCAACCGCAAAAACAGGCGATGATGTGATTAACCATTCAGTGTCTG 960

```

```

QY 961 TTTATCGCGGACAGATACTATCTGGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020
DB 961 TTTATCGCGGACAGATACTATCTGGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020
QY 1021 ACGCTTCCCGGTCTCAGCCCGGATTAACAGCCCGAGGTGTGAATCTGTGTTGAACGCTGG 1080
DB 1021 ACGCTTCCCGGTCTCAGCCCGGATTAACAGCCCGAGGTGTGAATCTGTGTTGAACGCTGG 1080
QY 1081 CGTGGCTTAAGGATTAACAGCCAGTGTGATTCAGTTCGTGCTTCAGACTTACAG 1140
DB 1081 CGTGGCTTAAGGATTAACAGCCAGTGTGATTCAGTTCGTGCTTCAGACTTACAG 1140
QY 1141 CAGATCGGTATTAACAGCCGCTGTCTGATTAATAGCCCGCGAGAGGTGAACCTGACC 1200
DB 1141 CAGATCGGTATTAACAGCCGCTGTCTGATTAATAGCCCGCGAGAGGTGAACCTGACC 1200
QY 1201 CTGGCAGATGTGAAGAGGAAATGCGCAGGCGATGTGTTGTTGCGAGTTTACGAA 1260
DB 1201 CTGGCAGATGTGAAGAGGAAATGCGCAGGCGATGTGTTGTTGCGAGTTTACGAA 1260
QY 1261 ATCGTAATGAAGACAGCATACCCGCTTGCACTTTGTA 1299
DB 1261 ATCGTAATGAAGACAGCATACCCGCTTGCACTTTGTA 1299

```

RESULT 9  
ADL16137  
ID ADL16137 standard; DNA; 1299 BP.

ADL16137;  
06-MAY-2004 (first entry)

Escherichia coli K-12 Appa phytase gene.

Appa phytase; modified phytase; K-12; animal feed; feed additive;  
phosphorus; phytate; myo-inositol hexakisphosphate; food production;  
wet grain milling; dental care product; bone resorption; osteoporosis;  
osteopathic; kidney stone; metal removal; waste treatment;  
cleaning product; rust removal product; gene; ds.

Escherichia coli; K-12.

Key Location/Qualifiers  
CDS 1..1299  
/\*tag= a  
/product= "Appa phytase"

NO2004015084-A2.

19-FEB-2004.

11-AUG-2003; 2003MO-US025058.

12-AUG-2002; 2002US-0403330P.

(GENEV ) GENENCOR INT INC.

Blattmann BO, Darzins A, Davis JM, Encell LP, Morrison TB;

Mrecho GT, Schellenberger V;

WPI; 2004-169498/16.

P-PSDS; ADL16129.

Novel recombinant phytase having modified phytase activity comprising  
modification of amino acid residues in mature Escherichia coli phytase  
designated EBC18B2, useful as feed additive.

Claim 19; Fig 21; 107pp; English.

The invention relates to recombinant modified Appa phytases having  
modified phytase activity. The Appa phytases are derived from a mature  
Escherichia coli Appa phytase designated EBC18B2 (see ADL16120), and

Query Match	99.6%	Score 1294.2	DB 12	Length 1299
Best Local Similarity	99.8%	Pred. NO. 0		
Matches 1296; Conservative	0	Mismatches	3	Indels 0; Gaps 0

OY	1	ANGAAGCATCTTAAATCCCATTTTATCTCTTGATTCGGTAAACCCGCGATCTGCA	60
Db	1	ANGAAGGCAATCTTAAATCCCATTTTATCTCTTGATTCGGTAAACCCGCGAATCTGCA	60
OY	61	TTGCGCTCAGATGAGCCGCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCAGTGCTATGGT	120
Db	61	TTGCGCTCAGATGAGCCGCGAGCTGAAAGTGTGTGATTTGTCAGTGCTATGGT	120
OY	121	GTGCGGCTCCAAACCAAGGCCAAGCAACTGATGACAGATGTCAACCCCAACGATGAGCA	180
Db	121	GTGCGGCTCCAAACCAAGGCCAAGCAACTGATGACAGATGTCAACCCCAACGATGAGCA	180
OY	181	ACCTGGCCGGTAAAACTGGGATTGGCTGACACCGCGCGGTGTGTAGCTAATTCGCTTATCTC	240
Db	181	ACCTGGCCGGTAAAACTGGGATTGGCTGACACCGCGCGGTGTGTAGCTAATTCGCTTATCTC	240
OY	241	GGAACATTACCAACGCGCAGCGGTCTGTAGCGCAACGATTTGCTGCGCAAAAAAGGCTGACCG	300
Db	241	GGAACATTACCAACGCGCAGCGGTCTGTAGCGCAACGATTTGCTGCGCAAAAAAGGCTGACCG	300
OY	301	CAGTCTGTGTCAGGTGCGCATTAATTGCTGATGTGCAGCAGCGTAAACCGTAAACAGCGCA	360
Db	301	CAGTCTGTGTCAGGTGCGCATTAATTGCTGATGTGCAGCAGCGTAAACCGTAAACAGCGCA	360
OY	361	GGCTTCGCGCGCGGGCTGGGCACTGACTGTGCAATAACGCTACATPACCAAGCGAGTACG	420
Db	361	GGCTTCGCGCGCGGGCTGGGCACTGACTGTGCAATAACGCTACATPACCAAGCGAGTACG	420
OY	421	TCCAGTCCCGATCCGTTATTTAAATCCCTTAAAAAATGCGCGTTTGCCAACTGTGATAACGCG	480
Db	421	TCCAGTCCCGATCCGTTATTTAAATCCCTTAAAAAATGCGCGTTTGCCAACTGTGATAACGCG	480
OY	481	AACGTACTGACGCGATCTCTCAGACAGGCGAGAGGGTCAATTGTGTAATTTACCGGGCAT	540
Db	481	AACGTACTGACGCGATCTCTCAGACAGGCGAGAGGGTCAATTGTGTAATTTACCGGGCAT	540

QY	541	GGGCAAGAGGGGGTTTCGGGAACGTGGAACGGGGCTTAAATTTTCCGCAATCAAACTGTGC	600
Db	541	CGGCAAAAGCGCGTTCCGGAACTGGAAACGGGTCCTAAATTTTCCGCAATCAAACTGTGC	600
QY	601	CTTAAACGTGAGAAACAGAGCGAAAGCTGTTCAATTAAACGACGACATTACCATCGAACTC	660
Db	601	CTTAAACGTGAGAAACAGAGCGAAAGCTGTTCAATTAAACGACGACATTACCATCGAACTC	660
QY	661	AAAGTGAGGGCCGACAAATGTCTCAATTAAACCGGTGGGTAAAGCTTCGATCAATGCTGACG	720
Db	661	AAAGTGAGGGCCGACAAATGTCTCAATTAAACCGGTGGGTAAAGCTTCGATCAATGCTGACG	720
QY	721	GAGATATTTCCTCTCAACCAAGACACAGGGAAATCCGGACCGGAGGTGGGGAAGATACCC	780
Db	721	GAGATATTTCCTCTCAACCAAGGGAAATCCGGAGCCGGGATGGGGAAGATACCC	780
QY	781	GATTCAACACAGTGGAAACACTTGTGTAAGTTTGCATTAACGCGCAATTTTATTTGTCACA	840
Db	781	GATTCAACACAGTGGAAACACTTGTGTAAGTTTGCATTAACGCGCAATTTTATTTGTCACA	840
QY	841	CGCAGCGCAAGAGTTGCCCGCAGCCGCGCACCCCGTTATTGATTTGATTCAGACAGCG	900
Db	841	CGCAGCGCAAGAGTTGCCCGCAGCCGCGCACCCCGTTATTGATTTGATTCAGACAGCG	900
QY	901	TTGAGCGCCCATCCACCCGCAAAAACAGCGGTATGAGTGACATTACCACTTCAGTGCCTG	960
Db	901	TTGAGCGCCCATCCACCCGCAAAAACAGCGGTATGAGTGACATTACCACTTCAGTGCCTG	960
QY	961	TTTATTCGCCGGAACAGATATCTAATCTGGCAAACTTCGCGCGCGCACTGAGACTCAACTGG	1020
Db	961	TTTATTCGCCGGAACAGATATCTAATCTGGCAAACTTCGCGCGCGCACTGAGACTCAACTGG	1020
QY	1021	ACGCTTCCCGTACGCCGATTAACACGCCGCAAGTGTGAACTGGTGTTTTGAACGCTGG	1080
Db	1021	ACGCTTCCCGTACGCCGATTAACACGCCGCAAGTGTGAACTGGTGTTTTGAACGCTGG	1080
QY	1081	CGTCGGCTAAGGATTAACAGCCAGAGGATTCAGGTTTCCTGTGCTTCCAGACTTTAACG	1140
Db	1081	CGTCGGCTAAGGATTAACAGCCAGTGTGATTCAGGTTTCCTGTGCTTCCAGACTTTAACG	1140
QY	1141	CAGATCCGTATPAAACGCCGCGTGTCAATTAAATACGCCGCCCGGAGAGGTGAACCTGACC	1200
Db	1141	CAGATCCGTATPAAACGCCGCGTGTCAATTAAATACGCCGCCCGGAGAGGTGAACCTGACC	1200
QY	1201	CTGGCAGAGTGTGAAGCGGAAATGCGCAGGCGATGTGTTCCGTGGCAGGTTTTCAGCAA	1260
Db	1201	CTGGCAGAGTGTGAAGCGGAAATGCGCAGGCGATGTGTTCCGTGGCAGGTTTTCAGCAA	1260
QY	1261	ATCGTAATGAAGACAGCATACCCGCTCAGATTGTAA	1299
Db	1261	ATCGTAATGAAGACAGCATACCCGCGCTCAGTTGTAA	1299

RESULT	10
AAD25463	
ID	AAD25463 standard; DNA; 1901 BP.
XX	
AC	AAD25463;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Escherichia coli appa phytase wild type DNA.
XX	
KM	Bacterial phytase; K12 appa phytase; protease stability; anaerobic;
KW	gastrointestinal; nutritional value; feed treatment process; therapy
KM	thermal tolerance; growth performance; alcoholic drink; biopulping;
KW	non-alcoholic drink; biobleaching; ds.
XX	
OS	Escherichia coli.
XX	
FH	Key
FT	Location/Qualifiers
CDS	188..1486
FT	/tag= a

/product="E. coli appa phytase protein"

WO200190333-A2.

29-NOV-2001.

24-MAY-2001; 2001WO-US017118.

25-MAY-2000; 2000US-00580515.

(DIVE-) DIVERSA CORP.

Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;

MPI; 2002-083108/11.

P-PSDB; AAE15807.

New bacterial phytase for e.g. improving the nutritional value of phytate-containing foodstuffs and subsequently improving the growth performance of an organism that consumes it, or in treating animal digestive systems.

Claim 42; Fig 7; 170pp; English.

The patent discloses recombinant bacterial phytase from *Escherichia coli* K12 appa phytase. The enzyme has phytase activity and improved thermal tolerance when compared with wild-type phytase. It has improved protease stability at low pH. The recombinant phytase is useful for improving the nutritional value of phytate-containing foodstuffs and subsequently improving the growth performance of an organism that consumes it, in treating animal digestive systems, in feed treatment processes and for in vitro purposes related to research, discovery and development. They are also used for generating recombinant digestive system life forms, for producing or manufacturing alcoholic and non-alcoholic drinks based on the use of moulds, grains and/or plants, in biopulping and bio-bleaching where a reduction in the use of environmentally harmful chemicals that are traditionally used in the pulp and paper industry is desired and in the reduction or possible elimination of the need for mineral supplements, enzymes or therapeutic drugs for animals from the daily feed. CC thus increasing the amount calories and nutrients present in the feed. CC The present sequence is a DNA encoding *E. coli* appa phytase wild type protein

Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;

Query Match 99.6%; Score 1293.2; DB 6; Length 1901;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCATTTTATCTCTTCTGATTCGTTAACCCGCAATCTGCA 60  
 188 ATGAAAGCCATCTTAATCCATTTTATCTCTTCTGATTCGTTAACCCGCAATCTGCA 247  
 61 TTGGCTCAGAGTACCGGAGCTGAAAGTGTGATGATGTCAGTCCGTCATGCT 120  
 248 TTGGCTCAGAGTACCGGAGCTGAAAGTGTGATGATGTCAGTCCGTCATGCT 307  
 121 GTGGTCTCTCAACCAAGCCAGCACTGATGATGATGATGATGATGATGATGATGAT 180  
 308 GTGGTCTCTCAACCAAGCCAGCACTGATGATGATGATGATGATGATGATGATGAT 367  
 181 ACCTGGCCGCTAAACTGGGTGCTGACACCGCGCGTGTGATGATGATGATGATGAT 240  
 368 ACCTGGCCGCTAAACTGGGTGCTGACACCGCGCGTGTGATGATGATGATGATGAT 427  
 241 GGACATTAACCAAGCCAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 428 GGACATTAACCAAGCCAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 487  
 301 CAGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 488 CAGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 547  
 361 GCCTTCCGCGCGGCTGACCTGATGATGATGATGATGATGATGATGATGATGATGAT 420

548 GCCTTCCGCGCGGCTGACCTGATGATGATGATGATGATGATGATGATGATGATGAT 607  
 421 TCCAGTCCGATCCGTTATTTATCTCTTCTGATTTGATGATGATGATGATGATGAT 480  
 608 TCCAGTCCGATCCGTTATTTATCTCTTCTGATTTGATGATGATGATGATGATGAT 667  
 481 AACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 668 AACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 727  
 541 CCGCAAAACGCGCTTTCGCAACTGTAACCGGCTTATTTTCCGCAATCAAACTTGTGC 600  
 728 CCGCAAAACGCGCTTTCGCAACTGTAACCGGCTTATTTTCCGCAATCAAACTTGTGC 787  
 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTTACCTGGAATC 660  
 788 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTTACCTGGAATC 847  
 661 AAGGTGAGGCGGACAAATGTTCAATTAACGAGGCTGTAAGCTGATCAATGTCAGC 720  
 848 AAGGTGAGGCGGACAAATGTTCAATTAACGAGGCTGTAAGCTGATCAATGTCAGC 907  
 721 GAGATATTTCTCTGCAACAGACAGGAAATGCGAGCCGCGGCTGAGGAAAGATCACC 780  
 908 GAGATATTTCTCTGCAACAGACAGGAAATGCGAGCCGCGGCTGAGGAAAGATCACC 967  
 781 GATTCAACCAAGTGAACACCTTGTCAATGTTGATTAACGAGCATTTTATTTGCTACAA 840  
 968 GATTCAACCAAGTGAACACCTTGTCAATGTTGATTAACGAGCATTTTATTTGCTACAA 1027  
 841 CCGACCCCAAGGTTTCCGCGAGCCGCGGCTGTAATGATTAATGATCAAGACGCG 900  
 1028 CCGACCCCAAGGTTTCCGCGAGCCGCGGCTGTAATGATTAATGATCAAGACGCG 1087  
 901 TTGACCCCAATCAACCGCAAAACAGGCTGATGATGATGATGATGATGATGATGATGAT 960  
 1088 TTGACCCCAATCAACCGCAAAACAGGCTGATGATGATGATGATGATGATGATGATGAT 1147  
 961 TTTATGCGCGGACAGATTAATCTGGAATCTGGCGGCGGCTGATGATGATGATGATGAT 1020  
 1148 TTTATGCGCGGACAGATTAATCTGGAATCTGGCGGCGGCTGATGATGATGATGATGAT 1207  
 1021 ACCGTTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 1208 ACCGTTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1267  
 1081 CGTGGCTAAAGCATTAACAGCCAGTGAATCAGGTTTCCGCTGCTTCCAGACTTTACAG 1140  
 1268 CGTGGCTAAAGCATTAACAGCCAGTGAATCAGGTTTCCGCTGCTTCCAGACTTTACAG 1227  
 1141 CAGATGCTGATTAACAGCCGCTGTCTTAATTAATACCGCGCGGAGAGTGAATCTGACC 1200  
 1328 CAGATGCTGATTAACAGCCGCTGTCTTAATTAATACCGCGCGGAGAGTGAATCTGACC 1387  
 1201 CTGGCAGATGATTAACAGGCAATTCGCGAGGCTGATGATGATGATGATGATGATGATGAT 1260  
 1388 CTGGCAGATGATTAACAGGCAATTCGCGAGGCTGATGATGATGATGATGATGATGATGAT 1447  
 1261 ATGTGAAATGAACAGCATTAACCGCTTGTGCAATTTGTA 1289  
 1448 ATGTGAAATGAACAGCATTAACCGCTTGTGCAATTTGTA 1486

RESULT 11  
 ADA19449  
 ID ADA19449 standard; DNA; 1901 BP.  
 AC ADA19449;  
 XX 20-NOV-2003 (first entry)  
 XX  
 DE E. coli K12 appa gene encoding phytase.



XX Phytase; ds; gene; phytase; appa gene; animal feed; inorganic phosphate;  
 KW digestion enhancement; transgenic; thermal tolerance; protease stability.  
 XX Escherichia coli; strain K12.  
 XX Key Location/Qualifiers  
 FT CDS 188..1486  
 FT /tag=a  
 FT /product="Phytase"  
 XX US2002136754-A1.  
 XX 26-SEP-2002.  
 XX 24-MAY-2001; 2001US-00866379.  
 XX 13-AUG-1997; 97US-00910798.  
 XX 01-MAR-1999; 99US-00259214.  
 XX 13-APR-1999; 99US-00291931.  
 XX 25-MAY-1999; 99US-00318528.  
 XX 25-MAY-2000; 2000US-00580515.  
 XX (SHOR/) SHORT J M.  
 XX (KRETT/) KRETT K A.  
 XX (GRAY/) GRAY K A.  
 XX (BART/) BARTON N R.  
 XX (GARR/) GARRETT J B.  
 XX (DONO/) O' DONOGHUE E.  
 XX (MATH/) MATHUR E J.  
 XX Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB;  
 XX O' Donoghue E, Mathur EJ;  
 XX MPI; 2003-040002/03.  
 XX P-PDB; ADA19450.  
 XX Isolated Escherichia coli polynucleotide encoding a modified phytase  
 PT enzyme, useful in the production of animal feed, for improving the  
 PT nutritional value of phytase-containing feedstuff and for enhancing  
 PT digestion in humans and animals.  
 XX Claim 42; Fig 7; 62pp; English.  
 XX The invention relates to an isolated Escherichia coli polynucleotide  
 CC encoding a phytase enzyme appearing as ADA19450 and having amino acids  
 CC modification from W68E, Q84W, A95P, K97C, S168E, R180Y, N226C or Y277D.  
 CC Also included the E. coli appa gene ADA19449 (or an oligonucleotide  
 CC derived from it) or its mutant sequence ADA19452, expression vectors,  
 CC host cells, a method of improving nutritional value of a phytase-  
 CC containing feedstuff by contacting the phytase-containing feedstuff with  
 CC a pure phytase enzyme mutated phytase (where the phytase enzyme catalyzes  
 CC the liberation of inorganic phosphate from the phytate in the phytate-  
 CC containing feedstuff), a method to produce an animal feed containing a  
 CC microbial phytase (comprising culturing the plant cell, plant part or  
 CC plant under conditions where the nucleotide sequence is expressed and  
 CC converting the plant cells, plants or plant into a composition for animal  
 CC feed), a feed composition for animals (comprising the plant seeds, plant  
 CC cells, plant parts or plants in admixture with a phytase-containing  
 CC feedstuff), a method to treat a human or an animal able to benefit from  
 CC digestive enhancement by the activity of an exogenous phytase enzymes  
 CC comprising administering to the human or animal the plant seed, plant  
 CC cell, plant parts or plants of a transgenic plant which is modified to  
 CC contain an expression system which expresses a nucleotide sequence  
 CC encoding a phytase enzyme, a transgenic non-human organism whose genome  
 CC comprising a heterologous nucleic acid sequence encoding a polypeptide  
 CC having phytase activity. The phytase enzyme is useful for improving the  
 CC nutritional value of phytase-containing feedstuff, in the production of  
 CC animal feed and for enhancing digestion in humans and animals. The  
 CC invented method improves thermal tolerance and protease stability. It  
 CC also improves the feeding value of phytate rich ingredients. The present  
 CC sequence represents the E. coli K12 appa gene encoding wild-type phytase.

SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;  
 Query Match 99.6%; Score 1293.2; DB 8; Length 1901;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 ATGAAGCCATCTTAATCCATTTTATCTTCTGATTCGGTTAACCCCGCAATCTGCA 60  
 DB 188 ATGAAGCCATCTTAATCCATTTTATCTTCTGATTCGGTTAACCCCGCAATCTGCA 247  
 QY 61 TTGCTCAGAGTGAAGCCGAGACTGAAGTGGAAAGTGGTATTCATGTCATCATGT 120  
 DB 248 TTGCTCAGAGTGAAGCCGAGACTGAAGTGGAAAGTGGTATTCATGTCATCATGT 307  
 QY 121 GTGCTGTCTCAACCAAGCCCACTGATGAGATGACACCCGAGCATGGCCA 180  
 DB 308 GTGCTGTCTCAACCAAGCCCACTGATGAGATGACACCCGAGCATGGCCA 367  
 QY 181 ACCTGCGCGTAAACTGGTGTGCTGACACCGCGGTGGTGAATATCGCTATCTC 240  
 DB 368 ACCTGCGCGTAAACTGGTGTGCTGACACCGCGGTGGTGAATATCGCTATCTC 427  
 QY 241 GGAATTAACAACGCGACGCTGTGTAGCCGAGATTCTGGCGCAAAAAGGCGTCCG 300  
 DB 428 GGAATTAACAACGCGACGCTGTGTAGCCGAGATTCTGGCGCAAAAAGGCGTCCG 487  
 QY 301 CAGTGTGTCAGGTGCGATTAATTTGCTGATGTCGACGCGTAAACCGTAAACAGCGCA 360  
 DB 488 CAGTGTGTCAGGTGCGATTAATTTGCTGATGTCGACGCGTAAACCGTAAACAGCGCA 547  
 QY 361 GCCTTGCGCGCGGCGTGGACCTGATCTGTCAATTAACGATACATACCAGGAGATACG 420  
 DB 548 GCCTTGCGCGCGGCGTGGACCTGATCTGTCAATTAACGATACATACCAGGAGATACG 607  
 QY 421 TCCAGTCCCGATCCGTTATTTATATCTCTTAAATAATGCGCTTTGGCACTGATACGCG 480  
 DB 608 TCCAGTCCCGATCCGTTATTTATATCTCTTAAATAATGCGCTTTGGCACTGATACGCG 667  
 QY 481 AACGTGACGTGACGCGATCCTGACGAGGCAAGAGGTCATTTGCTGATACCGGCGAT 540  
 DB 668 AACGTGACGTGACGCGATCCTGACGAGGCAAGAGGTCATTTGCTGATACCGGCGAT 727  
 QY 541 CGGCAAAAGCGCTTTGCGAACTGGAACGGGTCTTAAATTTTTCGCAATCAAACTTTGTC 600  
 DB 728 CGGCAAAAGCGCTTTGCGAACTGGAACGGGTCTTAAATTTTTCGCAATCAAACTTTGTC 787  
 QY 601 CTTAAACGTGAAACAGGACGAACGCTGTTATTAAGCAGGCTTAAACGATCGGAATC 660  
 DB 788 CTTAAACGTGAAACAGGACGAACGCTGTTATTAAGCAGGCTTAAACGATCGGAATC 847  
 QY 661 AAGGTGAGGCGGCAATGTCATTAAACGGGTGGTAAAGCTCGCATCAATGCTGACG 720  
 DB 848 AAGGTGAGGCGGCAATGTCATTAAACGGGTGGTAAAGCTCGCATCAATGCTGACG 907  
 QY 721 GAGATATTTCTCTGCAACAAACGAGGAATCCGAGCCGAGGTGGGAAAGATCAC 780  
 DB 908 GAGATATTTCTCTGCAACAAACGAGGAATCCGAGCCGAGGTGGGAAAGATCAC 967  
 QY 781 GATTACACCAAGTGAACACCTTGTGAATTTGCTAAGCCCATTTTATTTGCTACAA 840  
 DB 968 GATTACACCAAGTGAACACCTTGTGAATTTGCTAAGCCCATTTTATTTGCTACAA 1027  
 QY 841 CGCAGCGCGAGATTTGCGCGCGCGCACCCCGTTATTTGATTTGATCAAGACGCG 900  
 DB 1028 CGCAGCGCGAGATTTGCGCGCGCGCACCCCGTTATTTGATTTGATCAAGACGCG 1087  
 QY 901 TTGAGCGCCCATCAACCGCAAAACAGGCGTATGATGATCAATCAATCAATCAATGCTG 960  
 DB 1088 TTGAGCGCCCATCAACCGCAAAACAGGCGTATGATGATCAATCAATCAATCAATGCTG 1147  
 QY 961 TTTATGCGCGGACAGATCTAATCTGGCAATTCGCGGCGCATCGGAGCTCAACGCG 1020  
 DB 1148 TTTATGCGCGGACAGATCTAATCTGGCAATTCGCGGCGCATCGGAGCTCAACGCG 1207



QY 1021 AGCTTCCCGTACCGGATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1080  
DB 1208 AGCGTTCGCCGTACCGCGGATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1267  
QY 1081 CGTCCGCTAAGGATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1140  
DB 1268 CGTCCGCTAAGGATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1327  
QY 1141 CAGATCGTGAATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1200  
DB 1328 CAGATCGTGAATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1387  
QY 1201 CTGGCAGATGTGAAGGCGAAATGCGCAGGCGATGTGTTGTTGAGGTTTGAAC 1260  
DB 1388 CTGGCAGATGTGAAGGCGAAATGCGCAGGCGATGTGTTGTTGAGGTTTGAAC 1447  
QY 1261 ATCGTGAATGAACGAGGCGATGTGTTGTTGAGGTTTGAAC 1299  
DB 1448 ATCGTGAATGAACGAGGCGATGTGTTGTTGAGGTTTGAAC 1486

RESULT 12  
ADC87744  
ID ADC87744 standard; DNA; 1901 BP.  
XX  
AC ADC87744;  
XX  
DT 01-JAN-2004 (first entry)  
DE DNA encoding *Escherichia coli* appA phytase.  
XX  
XX Phytase; food supplement; enzyme delivery matrix; soybean meal;  
KM thermotolerance; thermostability; kernel; phytate;  
KM myo-inositol-hexaphosphate; inositol; inorganic phosphate;  
KM thermotolerant; feed value; digestion; gene; ds; appA.  
OS *Escherichia coli*.  
XX  
XX  
FH Key Location/Qualifiers  
FT 188..1486  
FT CDS  
FT /\*tag= a  
FT /product= "AppA phytase"  
FT /transl\_except= (pos:401..403,aa:Arg)

US2003103958-A1.  
XX  
XX  
PD 05-JUN-2003.  
XX  
PF 24-MAY-2002; 2002US-00156660.  
XX  
PR 13-AUG-1997; 97US-00910798.  
PR 01-MAR-1999; 99US-00259214.  
PR 13-APR-1999; 99US-00281931.  
PR 25-MAY-1999; 99US-00318528.  
PR 25-MAY-2000; 2000US-00580515.  
PR 24-MAY-2001; 2001US-00866379.  
XX  
XX (DIVE-) DIVERSA CORP.  
XX  
XX Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;  
PI Mathur BJ;  
XX  
XX WPI: 2003-787039/74.  
DR P-PSDB; ADC87745.  
DR  
DR  
PT New nucleic acid encoding a polypeptide having phytase activity, useful  
PT in improving the feeding value of phytate rich ingredients or as an aid  
PT in phytate digestion.  
XX  
XX  
PS Example 1, SEQ ID NO 3; 113pp; English.  
XX  
CC The invention discloses a new isolated or recombinant nucleic acid which

CC encodes a polypeptide having a phytase activity. Also claimed is a  
CC nucleic acid probe, an amplification primer sequence pair, an expression  
CC cassette comprising the nucleic acid, a vector comprising the nucleic  
CC acid, a transgenic non-human animal or plant, or its seed, comprising the  
CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of  
CC a phytase message in a cell, a heterodimer comprising the polypeptide and  
CC a second domain, an array comprising immobilised polypeptide or nucleic  
CC acid, a hydridoma comprising an antibody that specifically binds to the  
CC polypeptide, a food supplement for an animal, an edible enzyme delivery  
CC matrix, an edible pellet comprising a granule edible carrier and the  
CC polypeptide, a feed composition, a soybean meal, isolating or identifying  
CC the polypeptide, making an anti-phytase antibody, producing a recombinant  
CC polypeptide, determining whether a compound binds to the polypeptide,  
CC identifying a modulator whole cell engineering of new or modified  
CC phenotypes by using real-time metabolic flux analysis, increasing  
CC thermotolerance or thermostability of the phytase polypeptide, increasing  
CC the resistance of the phytase polypeptide to enzymatic inactivation in a  
CC digestive system and processing of corn and sorghum kernels. The phytase  
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to  
CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid  
CC is useful in improving the feeding value of phytate rich ingredients or  
CC as an aid in phytate digestion. The sequence presented is the DNA  
CC encoding the *Escherichia coli* appA phytase.  
XX  
SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;  
XX  
XX  
Query Match 99.6%; Score 1293.2; DB 10; Length 1901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ATGAAGCCATCTTATCCCATTTTATCTCTGATTCGTTAAACCCGCAATCTGA 60  
DB 188 ATGAAGCCATCTTATCCCATTTTATCTCTGATTCGTTAAACCCGCAATCTGA 247  
QY 61 TTGCTCAGAGTGAAGCCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120  
DB 248 TTGCTCAGAGTGAAGCCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 307  
QY 121 GTGCTGCTCAACCAAGCCAGCACTGATGAGATGATCAACCCAGCCGATGGCA 180  
DB 308 GTGCTGCTCAACCAAGCCAGCACTGATGAGATGATCAACCCAGCCGATGGCA 367  
QY 181 ACCGCGCCGTTAAACCTGGTGGCTGACACCGCGCGTGGTGAATCTGCTATCTC 240  
DB 368 ACCGCGCCGTTAAACCTGGTGGCTGACACCGCGCGTGGTGAATCTGCTATCTC 427  
QY 241 GGAATTATCCCAAGCCAGCTGTGTAAGCCGATGTGCGGAAAGGCTGCCG 300  
DB 428 GGAATTATCCCAAGCCAGCTGTGTAAGCCGATGTGCGGAAAGGCTGCCG 487  
QY 301 CAGCTGTGCAAGTCCGATTTATTTGCTGATGTGAAGGATACCCGTAAACAGGGGA 360  
DB 488 CAGCTGTGCAAGTCCGATTTATTTGCTGATGTGAAGGATACCCGTAAACAGGGGA 547  
QY 361 GCCTGCGCGCGCGCGCTGACCTGATGCAATACGTAATACCCAGGCGATAGC 420  
DB 548 GCCTGCGCGCGCGCGCTGACCTGATGCAATACGTAATACCCAGGCGATAGC 607  
QY 421 TCCAGTCCGATCGTATTTAAATCTCTAAACCTGCGCTTCCCACTGATTAACGCG 480  
DB 608 TCCAGTCCGATCGTATTTAAATCTCTAAACCTGCGCTTCCCACTGATTAACGCG 667  
QY 481 AACGTGATCAAGCGATCTCAAGCAGGCGAGAGGATCAATGCTGATTTTACCGGCA 540  
DB 668 AACGTGATCAAGCGATCTCAAGCAGGCGAGAGGATCAATGCTGATTTTACCGGCA 727  
QY 541 CCGCAAAACGCGCTTGGCGAATCGAAGCGGCTTAAATTTCCGCAATCAATCTGTC 600  
DB 728 CCGCAAAACGCGCTTGGCGAATCGAAGCGGCTTAAATTTCCGCAATCAATCTGTC 787  
QY 601 CTTAAAGTGAAGAAACAGAGCAAGAGCTGTTCATTAACGAGGATTAACATCGAATCTC 660  
DB 788 CTTAAAGTGAAGAAACAGAGCAAGAGCTGTTCATTAACGAGGATTAACATCGAATCTC 847

```

QY 661 AAGTGAGCGCCGACATATCTCATTAACCGGTGCGGTAGCCTCGCATCAATGCTGACG 720
    |||
Db 848 AAGTGAGCGCCGACATATCTCATTAACCGGTGCGGTAGCCTCGCATCAATGCTGACG 907
    |||
QY 721 GAGATATTTCTCTCTCAACAAGCAGAGGAATGCGGAGCGGGGTGGGAAAGATCAAC 780
    |||
Db 908 GAGATATTTCTCTCTCAACAAGCAGAGGAATGCGGAGCGGGGTGGGAAAGATCAAC 967
    |||
QY 781 GATTGACACAGATGGAACAACCTTGCTAAGTTTGATTAACGCGCAATTTTATTTGCTACAA 840
    |||
Db 968 GATTGACACAGATGGAACAACCTTGCTAAGTTTGATTAACGCGCAATTTTATTTGCTACAA 1027
    |||
QY 841 CGCAGCGCCAGAGGTTGCGCGGACCGCGCCACCCGTTTATGATTTGATCAAGACAGG 900
    |||
Db 1028 CGCAGCGCCAGAGGTTGCGCGGACCGCGCCACCCGTTTATGATTTGATCAAGACAGG 1087
    |||
QY 901 TTGAGCGCCCATCCACCGCAAAAACAGGGATATGATGATCAATTAACCCACTTCAAGTCTG 960
    |||
Db 1088 TTGAGCGCCCATCCACCGCAAAAACAGGGATATGATGATCAATTAACCCACTTCAAGTCTG 1147
    |||
QY 961 TTTATCGCCGACACAGATATCTAATCTGCAAAATCTCGGCGCGCATGAGCTCAACTGG 1020
    |||
Db 1148 TTTATCGCCGACACAGATATCTAATCTGCAAAATCTCGGCGCGCATGAGCTCAACTGG 1207
    |||
QY 1021 AGCGTTCCCGGTACAGCCGATTAACAGCCGCGCAGAGTGTGAATCTGTGTTTGAACGCTGG 1080
    |||
Db 1208 AGCGTTCCCGGTACAGCCGATTAACAGCCGCGCAGAGTGTGAATCTGTGTTTGAACGCTGG 1267
    |||
QY 1081 CGTGGCTAAGCGATTAACAGCCAGTGGATTTCAGTTTCGCTGCTTCCAGACTTTTACAG 1140
    |||
Db 1268 CGTGGCTAAGCGATTAACAGCCAGTGGATTTCAGTTTCGCTGCTTCCAGACTTTTACAG 1327
    |||
QY 1141 CAGATCGTGAATAAAACCGCGCTGTCAATTAATAACCGCCGCGGAGAGTGAATCTGACC 1200
    |||
Db 1328 CAGATCGTGAATAAAACCGCGCTGTCAATTAATAACCGCCGCGGAGAGTGAATCTGACC 1387
    |||
QY 1201 CTGGCAGATGTGAAGAAGGAAATGCGCAGGAGCATGTCTGTTGCGAGGTTTACGAA 1260
    |||
Db 1388 CTGGCAGATGTGAAGAAGGAAATGCGCAGGAGCATGTCTGTTGCGAGGTTTACGAA 1447
    |||
QY 1261 ATCGTGAATGAAGCAGCATATCCGCTTGAGATTTGTAA 1299
    |||
Db 1448 ATCGTGAATGAAGCAGCATATCCGCTTGAGATTTGTAA 1486
    |||

```

## RESULT 13

AD050297 standard; DNA; 1901 BP.

AD050297;

29-JUL-2004 (first entry)

*Escherichia coli* K12 *apra* phytase DNA.

*ApA* phytase; bacteric; thermal tolerance; protease stability; foodstuff; phytate; animal feed; fish feed; dough; baking; gene; ds.

*Escherichia coli*; K12.

Location/Qualifiers  
188..1486

FT /+tag= a  
FT /product= "ApA phytase"  
FT 401..403  
FT /+tag= b  
FT /note= "Encodes Arg"

US2004091968-A1.

13-MAY-2004.

XX

```

PF 20-JUN-2003; 2003US-00601319.
XX
PR 13-AUG-1997; 97US-00910798.
PR 01-MAR-1999; 99US-00259214.
PR 13-APR-1999; 99US-00291931.
PR 25-MAY-1999; 99US-00318528.
PR 25-MAY-2000; 2000US-00580515.
PR 24-MAY-2001; 2001US-00866379.
XX
PA (SHORT) SHORT J M.
PA (KREIZ) KREIZ K.
PA (GRAY) GRAY K A.
PA (BART) BARTON N R.
PA (GARF) GARRETT J B.
PA (ODON) O'DONOGHUE E.
PA (MATH) MATHER E J.
XX
PI Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
PI Mather EJ;
PI
DR MPI: 2004-374952/35.
DR P-PSDB: AD050298.
XX
PT Producing phytase, involves providing nucleic acid derived from bacteria
PT encoding polypeptide having phytase activity, and expressing nucleic acid
PT in yeast.
XX
PS Disclosure: SEQ ID NO 7; 74bp; English.
XX
CC The invention relates to a method for producing phytase that involves
CC providing a nucleic acid encoding phytase derived from a bacteria, and
CC expressing the nucleic acid in a yeast under conditions that allow
CC expression of the enzyme in the yeast. The invention also relates to
CC modified phytase enzyme which has improved thermal tolerance and protease
CC stability at low pH. The phytase enzyme can be used in foodstuffs to
CC improve the feeding value of phytate rich ingredients, and in diet of
CC numerous animals including mammals, fowls and fishes, commercially
CC significant mammals such as pigs, goats, laboratory rodents, commercially
CC significant avian species such as chicken, ducks, doves, parrot, etc.,
CC commercially farmed fish such as guppy, zebrafish, molly, swordtail,
CC etc., in dough making and baking, in dietary aids for animals. The method
CC provides easy manufacture of the active ingredient loaded biocompatible
CC composition, higher yields and loading efficiency. The phytase
CC incorporated in to the dietary aid is safe for animals. The present
CC sequence is Escherichia coli K12 apra phytase DNA.
XX
SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;
XX
Query Match 99.6%; Score 1293.2; DB 12; Length 1901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ATGAAAGCATCTTAATCCCATTTTATCTCTTGATTCGCTTAACCCCGAATCTGCA 60
    |||
Db 188 ATGAAAGCATCTTAATCCCATTTTATCTCTTGATTCGCTTAACCCCGAATCTGCA 247
    |||
QY 61 TTGCTCAGAGTGAGCGCGGAGCTGAAGTGAAGTGTGATTTGTCACTGATGCT 120
    |||
Db 248 TTGCTCAGAGTGAGCGCGGAGCTGAAGTGAAGTGTGATTTGTCACTGATGCT 307
    |||
QY 121 GTGCGTCTCAACCAAGGCGACGCAACTGATGAGATGTCAACCCGACGATGGCCA 180
    |||
Db 308 GTGCGTCTCAACCAAGGCGACGCAACTGATGAGATGTCAACCCGACGATGGCCA 367
    |||
QY 181 ACCTGCGCGGTAAACTGGGTTGGCTGACACCGCGCGGTGGAGCTATCGCTATTC 240
    |||
Db 368 ACCTGCGCGGTAAACTGGGTTGGCTGACACCGCGCGGTGGAGCTATTCGCTATTC 427
    |||
QY 241 GGACATTACCAACGACGAGCGTCTGGTAGCGGAGATTCTGGCGCAAAAAGGCTGCCG 300
    |||
Db 428 GGACATTACCAACGACGAGCGTCTGGTAGCGGAGATTCTGGCGCAAAAAGGCTGCCG 487
    |||
QY 301 CAGTCTGTCAAGTCCGATTAATGTGCTGATGTGACGAGCGTACCCGTAACAGGCGAA 360
    |||

```

Db	488	CAGCTGTGTGACAGTTCGGGATTTATGTGCTGATGTGACAGAGGTTACCGTAAACAGGGGAA	547
Qy	361	GCCTTTCGCGCGCGGACTGGCACTGACTGTGCATTAACCGTACATACCAGGCAAGATACG	420
Db	548	GCCTTTCGCGCGCGGACTGGCACTGACTGTGCATTAACCGTACATACCAGGCAAGATACG	607
Qy	421	TCCAGTCCCGCATCCGTTATTTAAATCTCTTAAACATGCGGTTTTCGCAACTGGATTAACGC	480
Db	608	TCCAGTCCCGCATCCGTTATTTAAATCTCTTAAACATGCGGTTTTCGCAACTGGATTAACGC	667
Qy	481	AACGTGACTGACGGATCTCTCAGACAGGGCAGAGGGTCAATTTGCTGACTTAAACCGGCAT	540
Db	668	AACGTGACTGACGGATCTCTCAGACAGGGCAGAGGGTCAATTTGCTGACTTAAACCGGCAT	727
Qy	541	CGGCAAAACGGCGTTTCGGGAACGTGAACGGGTGCTTAAATTTTCGCAATCAACTTGTGC	600
Db	728	CGGCAAAACGGCGTTTCGGGAACGTGAACGGGTGCTTAAATTTTCGCAATCAACTTGTGC	787
Qy	601	CTTAAACGTGGAACAGAGAGGAAAGCTGTTCAATTAACGACGACATTAACCATCGGAATC	660
Db	788	CTTAAACGTGGAACAGAGAGGAAAGCTGTTCAATTAACGACGACATTAACCATCGGAATC	847
Qy	661	AAGGTGAGCGCGGCAATATGTCATTAACGGGTGGGTAAACCTCGCATCAATCTGACG	720
Db	848	AAGGTGAGCGCGGCAATATGTCATTAACGGGTGGGTAAACCTCGCATCAATCTGACG	907
Qy	721	GAGATATTTTCTCTGCAACAAGACAAGGAAATGCCGAGACCGGGGTGGGGAAGATCAC	780
Db	908	GAGATATTTTCTCTGCAACAAGACAAGGAAATGCCGAGACCGGGGTGGGGAAGATCAC	967
Qy	781	GATTCAACAACAGTGAACAACCTTGCTTAAGTTTGCATTAACGGGCAATTTATTTTGTACAA	840
Db	968	GATTCAACAACAGTGAACAACCTTGCTTAAGTTTGCATTAACGGGCAATTTATTTTGTACAA	1027
Qy	841	CGCACGCGCAGAGGTGGCCCGGACCGCGGCAACCCGTTATTAAGATTGATTAAGACAGCG	900
Db	1028	CGCACGCGCAGAGGTGGCCCGGACCGCGGCAACCCGTTATTAAGATTGATTAAGACAGCG	1087
Qy	901	TTGACGCGCCCATCCACCGCAAAAACAGGCGATGAGTGTGACATTAACCCACTTCAGTGTCTG	960
Db	1088	TTGACGCGCCCATCCACCGCAAAAACAGGCGATGAGTGTGACATTAACCCACTTCAGTGTCTG	1147
Qy	961	TTTATGCGCGGACACGATACCTAATCTGCGCAATCTGCGCGCGCACCTGAGACTCAACTGG	1020
Db	1148	TTTATGCGCGGACACGATACCTAATCTGCGCAATCTGCGCGCGCACCTGAGACTCAACTGG	1207
Qy	1021	ACGCTTCCCGGTCACCGGATTAACAACGCCGCCAGGTGTGTGAACCTGTGTTGAACGCTGG	1080
Db	1208	ACGCTTCCCGGTCACCGGATTAACAACGCCGCCAGGTGTGTGAACCTGTGTTGAACGCTGG	1267
Qy	1081	CGTGGGCTAAAGGATTAACGACGAGTGTGATTCAGGTTTTCGCTGGTCTTCACGACTTAAAG	1140
Db	1268	CGTGGGCTAAAGGATTAACGACGAGTGTGATTCAGGTTTTCGCTGGTCTTCACGACTTAAAG	1327
Qy	1141	CAGATGCGGTATTAACCGCGCTGTCAATTAATACGCGCCCGGAGAGGTGAACCTGACC	1200
Db	1328	CAGATGCGGTATTAACCGCGCTGTCAATTAATACGCGCCCGGAGAGGTGAACCTGACC	1387
Qy	1201	CTGGCAGGATGTGAAGAGCGGAATGCGCAGGGCATGTGTCTGTGGCAGGTTTACGGA	1266
Db	1388	CTGGCAGGATGTGAAGAGCGGAATGCGCAGGGCATGTGTCTGTGGCAGGTTTACGGA	1447
Qy	1261	ATCGTAATGAAGACAGCATACCGGCTTGCAAGTTTGTA	1299
Db	1448	ATCGTAATGAAGACAGCATACCGGCTTGCAAGTTTGTA	1486
RESULT 14			
AD050295			
XX	AD050295	standard; DNA; 1901 BP.	
AC	AD050295;		

XX		29-JUL-2004	(first entry)	
DT				
XX	DE	Escherichia coli K12 apba phytase mutant DNA #1.		
XX	XX	Appa phytase; bacteria; thermal tolerance; protease stability; foodstuff;		
XX	KW	phyate; animal feed; fish feed; dough; baking; gene; mutant; ds.		
XX	OS	Escherichia coli; K12.		
OS	XX	Synthetic.		
XX	FT	Key	Location/Qualifiers	
FH	FT	mucation	/replace(389,T)	
FT	FT	mutation	/**tag= a replace(390,G) /*tag= b	
FT	XX			
XX	PV	US2004091968-A1.		
XX	PD	13-MAY-2004.		
PF	XX	20-JUN-2003; 2003US-00601319.		
PR	XX	13-AUG-1997; 97US-00910798.		
PR	XX	01-MAR-1999; 99US-00259214.		
PR	XX	13-APR-1999; 99US-00291931.		
PR	XX	25-MAY-1999; 99US-00318528.		
PR	XX	25-MAY-2000; 2000US-00580515.		
PR	XX	24-MAY-2001; 2001US-00866379.		
PA	XX	(SHOR/) SHORT J M.		
PA	XX	(KRET/) KRETTZ K.		
PA	XX	(GRAY/) GRAY K A.		
PA	XX	(BART/) BARTON N R.		
PA	XX	(GARF/) GARRETT J B.		
PA	XX	(ODON/) O'DONOUGHUE E.		
PA	XX	(MATH/) MATHER E J.		
PI	XX	Short JM, Kreutz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;		
PI	XX	Mather EJ;		
DR	XX	WPI; 2004-374952/35.		
PT	XX	Producing phytase, involves providing nucleic acid derived from bacteria		
PT	XX	encoding polypeptide having phytase activity, and expressing nucleic acid		
PT	XX	in yeast.		
PS	XX	Disclosure; SEQ ID NO 5; 74bp; English.		
CC	XX	The invention relates to a method for producing phytase that involves		
CC	XX	providing a nucleic acid encoding phytase derived from a bacteria, and		
CC	XX	expressing the nucleic acid in a yeast under conditions that allow		
CC	XX	expression of the enzyme in the yeast. The invention also relates to		
CC	XX	modified phytase enzyme which has improved thermal tolerance and protease		
CC	XX	stability at low pH. The phytase enzyme can be used in foodstuffs to		
CC	XX	improve the feeding value of phyrate rich ingredients, and in diet of		
CC	XX	numerous animals including mammals, fowls and fishes, commercially		
CC	XX	significant mammals such as pigs, goats, laboratory rodents, commercially		
CC	XX	significant avian species such as chicken, ducks, doves, parrot, etc.,		
CC	XX	commercially farmed fish such as guppy, zebrafish, molly, swordtail,		
CC	XX	etc., in dough making and baking, in dietary aids for animals. The method		
CC	XX	provides easy manufacture of the active ingredient loaded biocompatible		
CC	XX	composition, higher yields and loading efficiency. The phytase		
CC	XX	incorporated in to the dietary aid is safe for animals. The present		
CC	XX	sequence is Escherichia coli K12 apba phytase mutant DNA.		
SO	XX	Sequence 1901 BP; 475 A; 499 C; 499 G; 427 T; 0 U; 1 Other;		
QY	XX	Query Match      99.3%; Score 1290; DB 12; Length 1901; Best Local Similarity    99.5%; Pred. No. 0; Matches 1293; Conservative    0; Mismatches     6; Indels       0; Gaps       0		
QY	XX	1 ATGAAGAAGCATTTCATGCCATTTTTTACTCTTGATCGGTTAACCCGGCATCTCA 60		

```

Db      188 ATGAAAGGATCTTATCCCATTTTATCTCTTCGATTCGGTTAACCCCGCAATCTGCA 247
Qy      61 TTGCTGAGAGTAGAGCCGAGAGCTGAAGCTGGAAGTGTGTGATTTGTCACTGTCATGCT 120
Db      248 TTGCTGAGAGTAGAGCCGAGAGCTGAAGCTGGAAGTGTGTGATTTGTCACTGTCATGCT 307
Qy      121 GTGCGTGTCTCCAAACCAAGGCCCACTGATGATGAGATGTCAACCCGAGCATAGGCCA 180
Db      308 GTGCGTGTCTCCAAACCAAGGCCCACTGATGATGAGATGTCAACCCGAGCATAGGCCA 367
Qy      181 ACCTGGCCGGTAAACTGGGTTGGCTGACACCGCCGGTGTGAGCTAATGCTATCTC 240
Db      368 ACCTGGCCGGTAAACTGGGTTGGCTGAGCTGACACCGCCGGTGTGAGCTAATGCTATCTC 427
Qy      241 GGACATTAACCAAGCCAGGCTGTGTAGCCGACGATTTGCTGGCGAATAAGGGCTGCCG 300
Db      428 GGACATTAACCAAGCCAGGCTGTGTAGCCGACGATTTGCTGGCGAATAAGGGCTGCCG 487
Qy      301 CAGTCTGTGAGAGTGGCGATTTATGCTGATGTCGACGAGCCGTAACCCGTAACAGGCGAA 360
Db      488 CAGTCTGTGAGAGTGGCGATTTATGCTGATGTCGACGAGCCGTAACCCGTAACAGGCGAA 547
Qy      361 GCCTTCGCGCGGGCTGSCACTGACTGTGCAATTAACCGTAACCCGAGCAGATACG 420
Db      548 GCCTTCGCGCGGGCTGSCACTGACTGTGCAATTAACCGTAACCCGAGCAGATACG 607
Qy      421 TCCAGTCCCGATCCGTTATTTAATCCTCTAATAAAGCTGGCTTGGCCAACTGATACGG 480
Db      608 TCCAGTCCCGATCCGTTATTTAATCCTCTAATAAAGCTGGCTTGGCCAACTGATACGG 667
Qy      481 AACGTGACGAGCGCATCTCTGACGAGGAGGAGGAGGATGCAATTCGACTTTACCGGCGAT 540
Db      668 AACGTGACGAGCGCATCTCTGACGAGGAGGAGGAGGATGCAATTCGACTTTACCGGCGAT 727
Qy      541 CGGCAAAACGGCGTTTGGCAACTGGAACGGTGTCTTAATTTTCCGCAATCAAACTTGTGC 600
Db      728 CGGCAAAACGGCGTTTGGCAACTGGAACGGTGTCTTAATTTTCCGCAATCAAACTTGTGC 787
Qy      601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTAACATGGAATC 660
Db      788 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTAACATGGAATC 847
Qy      661 AAGGTAGAGGCGCAAGATGCTATTAACCGGTGCGGTAGAGCTCGCATCAATGCTGAG 720
Db      848 AAGGTAGAGGCGCAAGATGCTATTAACCGGTGCGGTAGAGCTCGCATCAATGCTGAG 907
Qy      721 GAGATATTTCTCTGCAACAGACAGGAAATGCCGAGCCGGGATGGGAAAGATCACC 780
Db      908 GAGATATTTCTCTGCAACAGACAGGAAATGCCGAGCCGGGATGGGAAAGATCACC 967
Qy      781 GATTCAACCAAGTGAACACCTTGTCTAATGTTGCAATTAACCGCAATTTATTTGCTACAA 840
Db      968 GATTCAACCAAGTGAACACCTTGTCTAATGTTGCAATTAACCGCAATTTATTTGCTACAA 1027
Qy      841 CGCAGCGCAGAGGTGGCGCGCAGCCGCGCCACCCGTTATTAGATTGATCAAGACAGC 900
Db      1028 CGCAGCGCAGAGGTGGCGCGCAGCCGCGCCACCCGTTATTAGATTGATCAAGACAGC 1087
Qy      901 TTGACGCCCATCAACCGCAAAAACAGCGTATGTTGATCAATTAACCACTTCAAGTCTG 960
Db      1088 TTGACGCCCATCAACCGCAAAAACAGCGTATGTTGATCAATTAACCACTTCAAGTCTG 1147
Qy      961 TTTATGCGCGGACACGATTAATCTGGAATCTGCGCGCGCACTGAGAGCTCACTG 1020
Db      1148 TTTATGCGCGGACACGATTAATCTGGAATCTGCGCGCGCACTGAGAGCTCACTG 1207
Qy      1021 AGCGTTCCGGGTAGCGCGATTAACAGCGCGCGAGTGTGAATGTTGAAACGCTGG 1080
Db      1208 AGCGTTCCGGGTAGCGCGATTAACAGCGCGCGAGTGTGAATGTTGAAACGCTGG 1267
Qy      1081 CGTCGGTTAAGCATTAACAGCATGTGATTCAAGTTTCTGCTGTTCTTCAGACTTTACAG 1140

```

```

Db      1268 CGTCGGTTAAGCATTAACAGCATGTGATTCAAGTTTCGCTGTTCCAGACTTTACAG 1327
Qy      1141 CAGATGGGTGATTAAGCGCGCTGTCAATTAATACGCCGCCCGGAGAGTGAATGAC 1200
Db      1328 CAGATGGGTGATTAAGCGCGCTGTCAATTAATACGCCGCCCGGAGAGTGAATGAC 1387
Qy      1201 CTGCGAGAGTGTGAAGACGAAATGCGACGAGCATGTGTTGTTGGCAGGTTTAACGCA 1260
Db      1388 CTGCGAGAGTGTGAAGACGAAATGCGACGAGCATGTGTTGTTGGCAGGTTTAACGCA 1447
Qy      1261 ATCGTAATGAAGCATACCGCATCCGCTTGCAGTTTGTAA 1299
Db      1448 ATCGTAATGAAGCATACCGCATCCGCTTGCAGTTTGTAA 1486

RESULT 15
AD050296
ID AD050296 standard; DNA; 1901 BP.
XX
XX AD050296;
AC
AC 29-JUN-2004 (first entry)
DT
DT 29-JUN-2004 (first entry)
XX
XX DE Escherichia coli K12 appA phytase mutant DNA #2.
XX
XX KW AppA phytase; bacteria; thermal tolerance; protease stability; foodstuff;
XX
XX KM phytase; animal feed; fish feed; dough; baking; gene; mutant; ds.
XX
XX OS Escherichia coli; K12.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX
XX FT mutation replace(437,C)
XX
XX FT mutation /*tag= a
XX
XX FT mutation replace(438,A)
XX
XX FT mutation /*tag= b
XX
XX FT mutation replace(439,A)
XX
XX FT mutation /*tag= c
XX
XX PN US2004091968-A1.
XX
XX PD 13-MAY-2004.
XX
XX PF 20-JUN-2003; 2003US-00601319.
XX
XX PR 13-AUG-1997; 97US-00910798.
XX
XX PR 01-MAR-1999; 99US-00259214.
XX
XX PR 13-APR-1999; 99US-00291931.
XX
XX PR 25-MAY-1999; 99US-00318528.
XX
XX PR 24-MAY-2000; 2000US-00580515.
XX
XX PR 24-MAY-2001; 2001US-00866379.
XX
XX PA (SHOR/) SHORT J M.
XX
XX PA (KRET/) KRETZ K.
XX
XX PA (GRAY/) GRAY K A.
XX
XX PA (BART/) BARTON N R.
XX
XX PA (GARR/) GARRETT J B.
XX
XX PA (ODON/) O'DONOGHUE E.
XX
XX PA (MATH/) MATHER E J.
XX
XX PI Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
XX
XX PI Mather EJ;
XX
XX DR MPI; 2004-374952/35.
XX
XX XX Producing phytase, involves providing nucleic acid derived from bacteria
XX
XX PT encoding polypeptide having phytase activity, and expressing nucleic acid
XX
XX PT in yeast.
XX
XX PS Disclosure; SEQ ID NO 6; 74pp; English.
XX
XX CC The invention relates to a method for producing phytase that involves
XX
XX CC providing a nucleic acid encoding phytase derived from a bacteria, and

```

expressing the nucleic acid in a yeast under conditions that allow expression of the enzyme in the yeast. The invention also relates to modified phytase enzyme which has improved thermal tolerance and protease stability at low pH. The phytase enzyme can be used in feedstuffs to improve the feeding value of phytate rich ingredients, and in diet of numerous animals including mammals, fowls and fishes, commercially. Significant mammals such as pigs, goats, laboratory rodents, commercially significant avian species such as chicken, ducks, doves, parrot, etc., commercially farmed fish such as guppy, zebrafish, molly, swordtail, etc., in dough making and baking, in dietary aids for animals. The method provides easy manufacture of the active ingredient loaded biocompatible composition, higher yields and loading efficiency. The phytase incorporated in to the dietary aid is safe for animals. The present sequence is Escherichia coli K12 appa phytase mutant DNA.

Sequence 1901 BP; 472 A; 498 C; 501 G; 429 T; 0 U; 1 Other;

Query Match 99.2%; Score 1288.4; DB 12; Length 1901;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1292; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy	1	ATGAAGCCATCTTAAATCCCATTTTATCTCTGATTCGGTAAACCCGCAATCTGCA	60
Db	188	ATGAAGCCATCTTAAATCCCATTTTATCTCTGATTCGGTAAACCCGCAATCTGCA	247
Qy	61	TTGCGTCAAGTAGAGCCGAGCTGAAAGTGTGATTTGTCATGTCATGTCATGTC	120
Db	248	TTGCGTCAAGTAGAGCCGAGCTGAAAGTGTGATTTGTCATGTCATGTCATGTC	307
Qy	121	GTTCGCTCTCAACCAAGGCGCAACCTGATCGAGATGTCACCCGAGCCGATGCGCA	180
Db	308	GTTCGCTCTCAACCAAGGCGCAACCTGATCGAGATGTCACCCGAGCCGATGCGCA	367
Qy	181	ACCTGCGCGGTAAATCTGAGTGTGCTGACACCGCGCGTGTGATGATTCGCTATCTC	240
Db	368	ACCTGCGCGGTAAATCTGAGTGTGCTGACACCGCGCGTGTGATGATTCGCTATCTC	427
Qy	241	GACATTATCAACGCGCGCTGTGTAGCCGAGCGAGTTCGTGCGAATAAGGCTGCGCG	300
Db	428	GACATTATCAACGCGCGCTGTGTAGCCGAGCGAGTTCGTGCGAATAAGGCTGCGCG	487
Qy	301	CAGTCGTGATGAGTGTGATTTATGTCGATGTGAGAGGTCACCGTAAACAGAGCGAA	360
Db	488	CAGTCGTGATGAGTGTGATTTATGTCGATGTGAGAGGTCACCGTAAACAGAGCGAA	547
Qy	361	GCCTTGCGCGCGCGGTGAGCACTGACTGTGCAATAACCGTACATACCCAGGAGATACG	420
Db	548	GCCTTGCGCGCGCGGTGAGCACTGACTGTGCAATAACCGTACATACCCAGGAGATACG	607
Qy	421	TCCAGTCCGATCCGTTATTTATCTCTTAAATACTGGCGTTTGCCAACTGGAATACGGC	480
Db	608	TCCAGTCCGATCCGTTATTTATCTCTTAAATACTGGCGTTTGCCAACTGGAATACGGC	667
Qy	481	AACGTGATGAGCGGATCTCAGAGGCGAGAGGTCATTTGTCGACTTTAACCGGGCAT	540
Db	668	AACGTGATGAGCGGATCTCAGAGGCGAGAGGTCATTTGTCGACTTTAACCGGGCAT	727
Qy	541	CGGCAACCGCGGTTTCCGAACTGAAACGGGTGCTTAATTTTCCGCAATCAACTGTGC	600
Db	728	CGGCAACCGCGGTTTCCGAACTGAAACGGGTGCTTAATTTTCCGCAATCAACTGTGC	787
Qy	601	CTTAAACGTGAGAAACAGAGCAAAAGCTGTCATTAAACGAGGCAATTAACATCGAACTC	660
Db	788	CTTAAACGTGAGAAACAGAGCAAAAGCTGTCATTAAACGAGGCAATTAACATCGAACTC	847
Qy	661	AAGGTGAGCGCGGCAATGTCATTTAAACCGGTGCGTAAAGCTCGCATGAATGTCAGC	720
Db	848	AAGGTGAGCGCGGCAATGTCATTTAAACCGGTGCGTAAAGCTCGCATGAATGTCAGC	907
Qy	721	GAGATATTTTCTCTGCAACAGCAACAGGAATGCGAGCGCGGGTGGGGAAGATCACCC	780
Db	908	GAGATATTTTCTCTGCAACAGCAACAGGAATGCGAGCGCGGGTGGGGAAGATCACCC	967

Qy	781	GATTACACAGAGGAAACACTTGTAGTTTGATTAACCGCAATTTTATTTGCTACAA	840
Db	968	GATTACACAGAGGAAACACTTGTAGTTTGATTAACCGCAATTTTATTTGCTACAA	1027
Qy	841	CGCAGCCAGAGGTTTCCGCGAGCCGCGCACCCGTTATTTAGTTTATCAAGACAGCG	900
Db	1028	CGCAGCCAGAGGTTTCCGCGAGCCGCGCACCCGTTATTTAGTTTATCAAGACAGCG	1087
Qy	901	TTGAGCCCGCATTCACCCGCAAAACAGGGTATGTTGTCATTAACCCCTTCAGTCGTC	960
Db	1088	TTGAGCCCGCATTCACCCGCAAAACAGGGTATGTTGTCATTAACCCCTTCAGTCGTC	1147
Qy	961	TTTATGCGCGGACACGATTAATCTGCAATCTCGGCGCGCATGAGCTCAACTGCG	1020
Db	1148	TTTATGCGCGGACACGATTAATCTGCAATCTCGGCGCGCATGAGCTCAACTGCG	1207
Qy	1021	ACGCTTCCGCGTACCGCGATTAACACCGCGCAGGTGTGAACTGCTTTTGAACGCTCG	1080
Db	1208	ACGCTTCCGCGTACCGCGATTAACACCGCGCAGGTGTGAACTGCTTTTGAACGCTCG	1267
Qy	1081	CGTGGCTAAGCATTAACAGCCAGTGTGATTCAGGTTTCCGCTTCCAGACTTTACAG	1140
Db	1268	CGTGGCTAAGCATTAACAGCCAGTGTGATTCAGGTTTCCGCTTCCAGACTTTACAG	1327
Qy	1141	CAGATGGGTATTAACGCGCTGTCTATTAAATACGCGCGCGGAGAGTGAACCTGACC	1200
Db	1328	CAGATGGGTATTAACGCGCTGTCTATTAAATACGCGCGCGGAGAGTGAACCTGACC	1387
Qy	1201	CTGGCAGATGTGAAGAGCGAAATGCGCAGAGGATGTGTTGTTGGCAGGTTTACGCA	1260
Db	1388	CTGGCAGATGTGAAGAGCGAAATGCGCAGAGGATGTGTTGTTGGCAGGTTTACGCA	1447
Qy	1261	ATGCTGATGAAGCAGCATATCCCGCTTGCAAGTTTGTAA	1239
Db	1448	ATGCTGATGAAGCAGCATATCCCGCTTGCAAGTTTGTAA	1486

Search completed: February 10, 2005, 09:41:14  
Job time : 889.787 secs

THIS PAGE BLANK (USPTO)



```

QY 241 AGGTTGTTCTGTGGGACAGTAGAAATTCAGAGAGGCGATGCGCTTTAGTGAACCTGATGG 300
DB 241 AGGGTGGTTCTGTGGGACAGTAGAAATTCAGAGAGGCGATGCGCTTTAGTGAACCTGATGG 300
QY 301 AAGCTACCCCAACAGACAGATGTTGTCAGTCAGGCGCAATCCGTTTCGATTTGATGGGCA 360
DB 301 AAGCTACCCCAACAGACAGATGTTGTCAGTCAGGCGCAATCCGTTTCGATTTGATGGGCA 360
QY 361 GCCGGACAGTGAAGACAGACACCTTACTGATTTGAGAGAGATGAAACAATGGCCAGC 420
DB 361 GCCGGACAGTGAAGACAGACACCTTACTGATTTGAGAGAGATGAAACAATGGCCAGC 420
QY 421 AGGATTTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCACTGATG 480
DB 421 AGGATTTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCACTGATG 480
QY 481 ACAAGCTGCCAAGAGACCA 500
DB 481 ACAAGCTGCCAAGAGACCA 500

```

```

RESULT 2
AX042372 20623 bp DNA linear PAT 23-NOV-2000
LOCUS Sequence 1 from Patent WO0064247.
DEFINITION AX042372.1 GI:11340990
ACCESSION AX042372.1
VERSION AX042372.1
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1. Forsberg, C.W., Golovan, S. and Phillips, J.P.
   Transgenic animals expressing salivary proteins
   Patent: WO 0064247-A 1 02-NOV-2000;
   UNIVERSITY OF GUELPH (CA)
JOURNAL

```

```

FEATURES
    source
        1..20623
            Location/Qualifiers
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="Tama2/APPA plasmid"

```

```

ORIGIN
Query Match 100.0%; Score 500; DB 6; Length 20623;
Best Local Similarity 100.0%; Pred. No. 1.1e-153;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TCGAGAGTATCTTTGTACAGTCGCTCCACAAAGGGGTACTGTTGCCACATAGAAG 60
DB 1 TCGAGAGTATCTTTGTACAGTCGCTCCACAAAGGGGTACTGTTGCCACATAGAAG 60
QY 61 ATCTAACTAATTAATTAATCCCTCAACCGCAATCTTTCAGTCACTAAGTTAGACGAT 120
DB 61 ATCTAACTAATTAATTAATCCCTCAACCGCAATCTTTCAGTCACTAAGTTAGACGAT 120
QY 121 TGTGTAACAAGTCTCCAAAGAGAGATACAGATGATGCGGTATAGGGTGAACCTGGCTG 180
DB 121 TGTGTAACAAGTCTCCAAAGAGAGATACAGATGATGCGGTATAGGGTGAACCTGGCTG 180
QY 181 CTGAGAGACACCTGATCTGACTAAGAGAGCCAGGCGTGAATGGATGGAGAT 240
DB 181 CTGAGAGACACCTGATCTGACTAAGAGAGCCAGGCGTGAATGGATGGAGAT 240
QY 241 AGGTTGTTCTGTGGGACAGTAGAAATTCAGAGAGGCGATGCGCTTTAGTGAACCTGATGG 300
DB 241 AGGTTGTTCTGTGGGACAGTAGAAATTCAGAGAGGCGATGCGCTTTAGTGAACCTGATGG 300
QY 301 AAGCTACCCCAACAGACAGATGTTGTCAGTCAGGCGCAATCCGTTTCGATTTGATGGGCA 360
DB 301 AAGCTACCCCAACAGACAGATGTTGTCAGTCAGGCGCAATCCGTTTCGATTTGATGGGCA 360

```

```

QY 361 GCCGACAGTGAAGACAGACACCTTACTGATTTGAGAGAGATGAGAAACAATGGCCAGC 420
DB 361 GCCGACAGTGAAGACAGACACCTTACTGATTTGAGAGAGATGAGAAACAATGGCCAGC 420
QY 421 AGGATTTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCACTGATG 480
DB 421 AGGATTTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCACTGATG 480
QY 481 ACAAGCTGCCAAGAGACCA 500
DB 481 ACAAGCTGCCAAGAGACCA 500

```

```

RESULT 3
AL732466 172535 bp DNA linear ROD 24-OCT-2002
LOCUS Mouse DNA sequence from clone RP23-149G14 on chromosome 2, complete
DEFINITION AL732466.
ACCESSION AL732466
VERSION AL732466.7 GI:24394996
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 172535)
REFERENCE
Direct Submission
Submitted (20-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:21621699.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Bm., EMBL, SW., SWISSPROT, Tr., TREMBL, Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-149G14 is from the RP23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.choi.oriy/bacpac/home.htm> VECTOR: pBACe3.6.

```

FEATURES
    source
        1..172535
            Location/Qualifiers
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /chromosome="2"
                /clone="RP23-149G14"
                /clone_lib="RP23-23"

```

```

ORIGIN
Query Match 100.0%; Score 500; DB 10; Length 172535;
Best Local Similarity 100.0%; Pred. No. 1.6e-153;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

QY      1 TCAGAGATATCTTTGTACGTGTGCTCCAAAGGGGTACTGTTCCCATATAGAAAG 60
Db      19120 TCAGAGATATCTTTGTACGTGTGCTCCAAAGGGGTACTGTTCCCATATAGAAAG 19179
QY      61 ATCTAAATCTATTAATTAATTCCTTCCACCGGCAATCTTTACGACCTAATAGTATGACGAT 120
Db      19180 ATCTAAATCTATTAATTAATTCCTTCCACCGGCAATCTTTACGACCTAATAGTATGACGAT 19239
QY      121 TGTGTAACAGATCTTCCAAAGAGAGATACAGATGATGCGGTATAGGGTGAACCTGCTG 180
Db      19240 TGTGTAACAGATCTTCCAAAGAGAGATACAGATGATGCGGTATAGGGTGAACCTGCTG 19299
QY      181 CTGAGAGACACCTGATCTGATTAAGAGAGCCGCGTGTATGATGATGTTGTGAGT 240
Db      19300 CTGAGAGACACCTGATCTGATTAAGAGAGCCGCGTGTATGATGATGTTGTGAGT 19359
QY      241 AGGTGTGTTCTGTGGGACAGTATGAAAATCGAGAGGATGTCGGTGTATGAACTGATG 300
Db      19360 AGGTGTGTTCTGTGGGACAGTATGAAAATCGAGAGGATGTCGGTGTATGAACTGATG 19419
QY      301 AAGCTACCCCAACGACAGAGATGTCAGTCAAGGCCATCCGTTGCGAGTTTGAAGGCA 360
Db      19420 AAGCTACCCCAACGACAGAGATGTCAGTCAAGGCCATCCGTTGCGAGTTTGAAGGCA 19479
QY      361 GCCGACAGTGAAGACACACCTACTACTGTTGAGAGAGATGAGAACATAGCCAGC 420
Db      19480 GCCGACAGTGAAGACACACCTACTACTGTTGAGAGAGATGAGAACATAGCCAGC 19539
QY      421 AGGATTTGAGAGACCTTGAAGGCGCAAGGCCCTTAACACACACCTACCTCACTTG 480
Db      19540 AGGATTTGAGAGACCTTGAAGGCGCAAGGCCCTTAACACACACCTACCTCACTTG 19599
QY      481 ACAAGCTGCAAGAACCA 500
Db      19600 ACAAGCTGCAAGAACCA 19619

RESULT 4
AC107726 214345 bp DNA linear HTG 03-MAR-2003
LOCUS Mus musculus clone RP24-85E12, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
AC107726 GI:28613318
VERSION HTG: HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214345)
2 (bases 1 to 214345)
Mus musculus, clone RP24-85E12
Unpublished
2 (bases 1 to 214345)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dekrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Gardner, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M.,
McWan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Reita, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 214345)
Birren, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dekrellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Nguyen, C., O'Connor, T., O'Donnell, P.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Reita, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Wassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2003 this sequence version replaced gi:28173217.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20103
Center clone name: 85 E 12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 21253 bases at least Q40
Consensus quality: 21324 bases at least Q30
Consensus quality: 21351 bases at least Q20
Insert size: 19400; agarose-fp
Insert size: 213645; sum-of-coverage
Quality coverage: 13.0 in Q20 bases; agarose-fp
Quality coverage: 11.8 in Q20 bases; sum-of-coverage
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 8 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
64960: contig of 64960 bp in length
64961: gap of 100 bp
65060: contig of 3215 bp in length
65061: gap of 100 bp
68275: gap of 100 bp
68276: gap of 100 bp
68376: contig of 6153 bp in length
68376: gap of 100 bp
74568: contig of 6153 bp in length
74568: gap of 100 bp
74669: contig of 6353 bp in length
81022: gap of 100 bp
81022: contig of 10329 bp in length
81132: gap of 100 bp
81132: contig of 10329 bp in length
91451: gap of 100 bp
91451: contig of 10329 bp in length

```

FEATURES

\* 91551 118992: contig of 27442 bp in length

\* 118993 119092: gap of 100 bp

\* 119093 165215: contig of 46123 bp in length

\* 165216 165315: gap of 100 bp

\* 165316 214345: contig of 49030 bp in length.

LOCATION/Qualifiers

SOURCE

1. 214345

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP24-85E12"

/clone\_lib="RP24 Male Mouse BAC"

1. 64960

/note="assembly\_fragment"

misc\_feature

/note="assembly\_fragment"

vector\_end:SP6

vector\_side:left"

65061. 68275

/note="assembly\_fragment"

68376. 74568

/note="assembly\_fragment"

74669. 81021

/note="assembly\_fragment"

81122. 91450

/note="assembly\_fragment"

91551. 118992

/note="assembly\_fragment"

119093. 165215

/note="assembly\_fragment"

165316. 214345

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

## ORIGIN

Query Match 100.0%; Score 500; DB 2; Length 214345;

Best local Similarity 100.0%; Pred. No. 1.7e-153;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCGAGAGTATCTTTGTGACGTGTGCTCCCAACAAAGGGGTACTGTTCCCACTAGAAAG 60

49912 TCGAGAGTATCTTTGTGACGTGTGCTCCCAACAAAGGGGTACTGTTCCCACTAGAAAG 49971

61 ATCTAATCAATTAATTAATTAATCCCTCAACCGGCAATCTTTCAGTCATTAAGTTAGACAGAT 120

49972 ATCTAATCAATTAATTAATTAATCCCTCAACCGGCAATCTTTCAGTCATTAAGTTAGACAGAT 50031

121 TGTGAACAAGTCTCCAAAGAGAGATACAGATGAGTGGTATAGAGTGGACCTGCTG 180

50032 TGTGAACAAGTCTCCAAAGAGAGATACAGATGAGTGGTATAGAGTGGACCTGCTG 50091

181 CTGAGAGACACTGTCATCTGACTTAAGAGAGCCGAGTGTAGTTGATGATGTTGGAGT 240

50092 CTGAGAGACACTGTCATCTGACTTAAGAGAGCCGAGTGTAGTTGATGATGTTGGAGT 50151

241 AGGGTGGTCTGGGGACAGTAGAAAATCGAGAGGATGCGGCTTAGTGAAGTGAATG 300

50152 AGGGTGGTCTGGGGACAGTAGAAAATCGAGAGGATGCGGCTTAGTGAAGTGAATG 50211

301 AAGCTACCCCAACGACAGAGATGTGACGTCAAGGCCAATCCGTTTCGAGTTTGAAGGGCA 360

50212 AAGCTACCCCAACGACAGAGATGTGACGTCAAGGCCAATCCGTTTCGAGTTTGAAGGGCA 50271

361 GCGGACAGTGAACAGACACACTTACTGAGTTGAGAGAGATGAGAAACAAAGGGCCAGC 420

50272 GCGGACAGTGAACAGACACACTTACTGAGTTGAGAGAGATGAGAAACAAAGGGCCAGC 50331

421 AGGATTTGAGAGACCTGACAGGGGCAAGGCCCTTAACACACACACCTACCACTCACTTG 480

50332 AGGATTTGAGAGACCTGACAGGGGCAAGGCCCTTAACACACACACCTACCACTCACTTG 50391

481 ACAAAAGCTGCCAAAGACCAA 500

50392 ACAAAAGCTGCCAAAGACCAA 50411

RESULT 5

AC103277/c

LOCUS

DEFINITION

AC103277 236802 bp DNA linear HTG 13-MAY-2003

Rattus norvegicus clone CH230-235gl, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 2 unordered pieces.

AC103277.6 GI:30579004

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 236802)

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D.,

Anyalebechi,Y., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryan,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ciesar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Dayala,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Derram,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Hayes,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,W.,

Georgescu,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M.,

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlik,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,

Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorenshewa,L., Louissege,H., Lozano,R.O., Lu,X., Ma,J.,

Maheshwari,M., Mahindratne,M., Mahmood,M., Malloy,K., Mangum,A.,

Manungu,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Manthey,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Murphy,M.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Natr,L.,

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

Nwackemele,O., Okwionu,G., Olariupona,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,

Plapper,F., Polindexter,A., Popovic,D., Primus,B., Pu,L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M., Reigh,R.,

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.,

Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,

Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajic,D.,

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

Steinle,M., Strong,R., Sutton,A., Svetek,A., Tabors,P., Taylor,C.,

Taylor,T., Thomas,N., Thomas,S., Tinney,A., Trefos,Z., Umanat,C.,

Vales,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,

Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

Williams,G., Willson,R., Wleczek,R., Wooden,H., Wotley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., Von

Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 236802)

Worley,K.C.

Direct Submission

Submitted (24-NOV-2001)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA



```

/translation="MADEKPKGKVTENNNDHNLKVAGDGSVVOFKIKRHTPLSKLM
KAYCERGLSMKQIFRFDPGQINETDPAOLEMDEDDTIVFOQQTGTV"
3'UTR
/gene="SMT3"
/contig="putative"
900..905
polyA_signal
/gene="SMT3"
/contig="putative"
915
polyA_site
/gene="SMT3"
/contig="putative"

ORIGIN
Query Match      19.3%; Score 96.6; DB 10; Length 915;
Best Local Similarity 63.7%; Pred. No. 2.8e-20;
Matches 179; Conservative 0; Mismatches 99; Indels 3; Gaps 2;

QY 221 TTAGTTGATGGTGGTGAAGGAGTCTGTCTGTGGGACAGTAAATCGAGAGCATGT 280
DB 87 TAAATTTGAAGGTGGCGGGGAGAGATGTTCTGTGGTGGGATTTAAGATTAGAGCATAC 146
QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAAGCAGAGATTGTCTAGGCCAAT 339
DB 147 ACCACTTAGTAAACTAATGAAAGCCATATTGTGACAGACAGGATTTGTCAATGAGCAGAT 206
QY 340 CCGTTTCGAGTTTGAAGGCGGACCGGACAGTGAACAGACACCTACTCAGTTGGAGGA 399
DB 207 CAGATTCCCGTTTGAAGCAGCAATCAATGAAAGACACCTGCAAGTTGGAAT 266
QY 400 AGGATGAGAACAATGCGCAGAGGATTTAGAGACCTTCAAGCGCCCAACAC 459
DB 267 GGAGATGAAGATACATTTGATGTG--TTCCAGACACAGACAGAGGTCTTACTTAAATA 324
QY 460 ACACACCTACCACTCTTGAACAAAGTCCCAAGACCA 500
DB 325 GGGACCTGCTACTTACTCTCAGAAATTTGTTATAGACCA 365

RESULT 7
AC119067 193199 bp DNA linear HTG 14-JAN-2003
LOCUS Canis familiaris clone RP81-23M24, WORKING DRAFT SEQUENCE, 11
DEFINITION
ordered pieces.
AC119067
VERSION AC119067.4 GI:27733945
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 193199)
Ahner, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Boufield, G.G., Brinkley, C., Brooks, S.,
Carigga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
Laric, P., Lee-Lin, S.-O., Legaspi, R., Maduro, O.L., McDowell, J.,
Margulies, E.H., Masello, C., Maerker, B., McDowell, J.,
Peduligan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
Reddix-Dugue, N., Schandler, K., Schuler, M.G., Simon, C.,
Stancir, P., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,
Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 193199)
Green, E.D.
TITLE Direct Submission
AUTHORS Submitted (24-APR-2002) NIH Intramural Sequencing Center, 8717
JOURNAL Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE 3 (bases 1 to 193199)
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

```

## COMMENT

On Jan 14, 2003 this sequence version replaced gi:24308653.

----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: nisc\_zoo@nhgri.nih.gov  
----- Project Information  
Center project name: dga  
Center clone name: 023M24

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

## ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 19149 bases at least Q40  
Consensus quality: 191542 bases at least Q30  
Consensus quality: 191909 bases at least Q20  
Insert size: 188000; agarose-fp  
Insert size: 192199; sum-of-contigs  
Quality coverage: 10.74x in Q20 bases; agarose-fp  
Quality coverage: 10.50x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

```

1 20054: contig of 20054 bp in length
* 20055 20154: gap of unknown length
* 20155 23880: contig of 3726 bp in length
* 23881 23980: gap of unknown length
* 23981 29813: contig of 5833 bp in length
* 29814 29913: gap of unknown length
* 29914 60037: contig of 30124 bp in length
* 60038 72455: gap of unknown length
* 72456 72555: gap of unknown length
* 72556 74766: contig of 2211 bp in length
* 74767 74866: gap of unknown length
* 74867 126334: contig of 51468 bp in length
* 126335 126434: gap of unknown length
* 126435 140029: contig of 13595 bp in length
* 140030 140129: gap of unknown length
* 140130 144562: contig of 4433 bp in length
* 144563 144662: gap of unknown length
* 144663 183707: contig of 39045 bp in length
* 183708 183808: gap of unknown length
* 183808 193199: contig of 9392 bp in length.

```

## FEATURES

## source

1..193199

/organism="Canis familiaris"

/mol\_type="genomic DNA"

/db\_xref="taxon:9615"

/clone\_lib="RP81-23M24"

/clone\_lib="RP81"

1..20054

## misc\_feature

/note="assembly\_fragment"

clone\_end:Sp6  
vector\_side:left"  
20155..23880  
/note="assembly\_fragment"  
23981..29813  
/note="assembly\_fragment"  
29914..60037  
/note="assembly\_fragment"  
60138..72455  
/note="assembly\_fragment"  
72556..74766  
/note="assembly\_fragment"  
74867..126334  
/note="assembly\_fragment"  
125898..193199  
/note="clone overlaps with GenBank Accession Number  
AC120521 clone RP81-392C9 (center project name dgb)"  
126435..140029  
/note="assembly\_fragment"  
140130..144562  
/note="assembly\_fragment"  
144663..183707  
/note="assembly\_fragment"  
183808..193199  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

## ORIGIN

Query Match 18.5%; Score 92.4; DB 2; Length 193199;  
Best Local Similarity 65.9%; Pred. No. 1.7e-18;  
Matches 149; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 206 AGAAGACCAGGCTGTAGTGAATGCTGAGTGAAGTGGTCTCTGGGACAGTAGAA 265  
DB 58554 AGAAGACGATCATATTAATCTGAAGTGGGGGCGAGATGTTCTATGTGACGCTTA 58613  
QY 266 AATCGAGAGCGATGCGCTTAACTGATGAATGATGAAG-CTACCCCAAAGACAGAGATT 324  
DB 58614 GATTAAAGAGCATACCATTAATTAATTAATTAAGCTACTGTGAACGACAGCGGTTT 58673  
QY 325 GTCAAGTACGCGCATCCGTTTCCAGTTTGAATGGGCGACCGGACAGTGAAGACAGCACC 384  
DB 58674 GTCAATGAGGACAGTTCGATTCGGGGTGAATGGGCAACCAATCAATAAACAGACACC 58733  
QY 385 TACTCAGTTGAGAGATGAGAAACAATGCGCCAGCAGGATTGAG 430  
DB 58734 TGCAAGCTGGAATGAGAAAGAAAGACAACTGTGTTCAG 58779

RESULT 8  
PTU89438 717 bp mRNA linear VRL 10-JAN-2000  
LOCUS PTU89438  
DEFINITION Pestivirus type 1 nonstructural protein p125-2 (p125) mRNA, partial  
cda.  
ACCESSION U89438  
VERSION U89438.1 GI:2707596  
KEYWORDS Bovine viral diarrhea virus 1 (BVDV-1)  
SOURCE Bovine viral diarrhea virus 1  
ORGANISM Pestivirus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Pestivirus.  
REFERENCE 1 (bases 1 to 717)  
AUTHORS Qi, F., Ridpath, J.F. and Berry, E.S.  
TITLE Insertion of a bovine SM7B gene in NS4B and duplication of NS3 in  
a bovine viral diarrhea virus genome correlate with the  
cytopathogenicity of the virus  
JOURNAL JOURNAL  
MEDLINE 99049626  
PUBMED 9833880  
REFERENCE 2 (bases 1 to 717)  
AUTHORS Qi, F. and Berry, E.S.  
TITLE Direct Submission

JOURNAL Submitted (12-FEB-1997) Oral Biology, University of Alabama at  
Birmingham, 250 LHRB 1919 7th Ave. S, Birmingham, AL 35294, USA  
FEATURES  
source  
location/Qualifiers  
1..717  
/organism="Bovine viral diarrhea virus 1"

gene  
CDS  
1..>717  
/mol\_type="mRNA"  
/strain="SNC"  
/db\_xref="taxon:11099"  
/note="p125 region"  
1..>717  
/gene="p125"  
1..>717  
/gene="p125"  
/note="the 3' region of the p125 protein is duplicated in  
this strain. Therefore, U89438 is the first copy, and the  
3' region of the U89438 sequence contains the second copy"

## ORIGIN

Query Match 18.1%; Score 90.4; DB 14; Length 717;  
Best Local Similarity 70.3%; Pred. No. 3.1e-18;  
Matches 135; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 221 TTAGTGAATGCTGCTGAGTGAAGTGGTCTCTGAGGACAGTGAATGAGAGGATG 280  
DB 222 TAATTTGAAGTGGCCGGGAGAGATGTTCTGAGCAGTTTAAGATTAAAGGCAATAC 281  
QY 281 GCCGTTTGAAGTGAATGAGTGAAG-CTACCCCAAAGACAGAGATTGTCACTCAGCCCAT 339  
DB 282 ACCACTTAAGTAATTAATTAAGAACCTATGTGAACAGCAGGCTTGTCAATGAGGAGAT 341  
QY 340 CCGTTTGAAGTGAATGAGGCGGACCGGACAGTGAACAGACACCTACTCAATTTGAGAGA 399  
DB 342 CAGATTCCGATTTTGAAGCGGCGGACCAATCAATGAAGACAGACACCTGACAGTTGAGAT 401  
QY 400 AGGATGAGAA 411  
DB 402 GGAAGATGAGAA 413

RESULT 9  
BTU89439 1151 bp mRNA linear MAM 10-JAN-2000  
LOCUS BTU89439  
DEFINITION Bos taurus ubiquitin-like protein mRNA, complete cds.  
ACCESSION U89439  
VERSION U89439.1 GI:1888535  
KEYWORDS Bos taurus (cow)  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1151)  
AUTHORS Qi, F., Ridpath, J.F. and Berry, E.S.  
TITLE Insertion of a bovine SM7B gene in NS4B and duplication of NS3 in  
cytopathogenicity of the virus  
JOURNAL JOURNAL  
MEDLINE 99049626  
PUBMED 9833880  
REFERENCE 2 (bases 1 to 1151)  
AUTHORS Qi, F. and Berry, E.S.  
TITLE Direct Submission  
JOURNAL Submitted (12-FEB-1997) Oral Biology, University of Alabama at  
Birmingham, 250 LHRB 1919 7th Ave. S, Birmingham, AL 35294, USA

```

FEATURES
  Source
    Location/Qualifiers
      1..1151
        /organism="Bos taurus"
        /mol_type="mRNA"
        /db_xref="taxon:9913"
        /tissue_type="endometrium"
        /size="548..835"
        /function="unknown"
        /codon_start=1
        /product="ubiquitin-like protein"
        /protein_id="BAB49682.1"
        /db_xref="GI:1888536"
        /translation="MADEPKKRGVTENNNDHNLKVAGDGSVVOFKIKRHPISKLM
        KAYCRQGLSHMQILFRPDGQIFINTDTPPAQLEMEDEDTIVFQQTGGYV"

ORIGIN
  Query Match      18.1%; Score 90.4; DB 4; Length 1151;
  Best Local Similarity 70.3%; Pred. No. 3.3e-18;
  Matches 135; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 221 TTATGTAAGTGTGTGAGTACGAGGTGTTCTGTGGACAGTGAAGAAATGAGAGGCAATCT 280
DB 601 TATTTGAAGTGTGGCGGCGGACAGATGTTCTGTGTGACGATTAAAGATTAAGACATAC 660
QY 281 GGCGTTAGTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTGAGGCCAAT 339
DB 661 ACCACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 720
QY 340 CCGTTTCGAGTTGATGAGCGGACGCGGACAGTGAAGACAGACACCTACTCACTGAGGAGA 399
DB 721 CAGATTCGAGTTGATGAGCGGACGCGGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780
QY 400 AGGATGAGAGCA 411
DB 781 GGAAATGAAGA 792

RESULT 10
LOCUS AC127104
DEFINITION Rattus norvegicus clone CH230-128A21, *** SEQUENCING IN PROGRESS
ACCESSION AC127104.3 GI:23264301
VERSION HTG; HTGS_PHASE1; HTGS_DNAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
  1 (bases 1 to 211409)
  Wuzny,D,Matte, Weizker,M, Lee, Abramson, S., Adams, C., Alder, J.,
  Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
  Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
  Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
  Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
  Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
  Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Center, A.,
  Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
  Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
  Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
  Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
  Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
  Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
  Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
  Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
  Gibragegisi, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
  Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K.,
  Harvey, Y., Hayak, P., Hawes, A., Henderson, N., Hernandez, J.,
  Hernandez, R., Hines, S., Hladin, S. L., Hodgson, A., Hognes, M.,
  Hollins, B., Howells, S., Huyl, S., Hume, J., Idelbr, D., Jackson, A.,
  Jackson, L., Jacob, L., Jiang, S., Johnson, B., Johnson, R., Jolivet, A.,
  Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

```

```

TITLE
JOURNAL
AUTHORS
REFERENCE
JOURNAL
COMMENT
  Unpublished
  2 (bases 1 to 211409)
  Worley, K. C.
  Direct Submission
  Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 211409)
  Rat Genome Sequencing Consortium.
  Direct Submission
  Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Sep 21, 2002 this sequence version replaced gi:21203531.
  The sequence in this assembly is a combination of BAC based reads
  and whole genome shotgun sequencing reads assembled using Atlas
  (http://www.bgsc.bcm.tmc.edu/projects/rat/). As a result, the
  sequence may extend beyond the ends of the clone and there may be
  contigs that consist entirely of whole genome shotgun sequence
  reads. Both end sequences and whole genome shotgun sequence only
  contigs will be indicated in the feature table.
  ----- Genome Center
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.bgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  ----- Project Information
  Center project name: GNMJX
  Center clone name: CH230-128A21
  ----- Summary Statistics
  Assembly program: Phrap; version 0.990329
  Consensus quality: 187810 bases at least Q40
  Consensus quality: 190349 bases at least Q20
  Consensus quality: 191813 bases at least Q20
  Estimated insert size: 209849; sum-of-contigs estimation
  Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

```

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_dirct\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_dirct_data.html)).  
 NOTE: This is a "working draft" sequence. It currently  
 consists of 3 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence

```

* as soon as it is available and the accession number will
* be preserved.
1 208510: contig of 208510 bp in length
* 208511 208610: gap of unknown length
* 208611 209996: contig of 1386 bp in length
* 209997 210096: gap of unknown length
* 210097 211409: contig of 1313 bp in length.
FEATURES
source
1. 211409
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-128A21"
892..1776
/notes="clone_boundary"
clone_end:17
site:ECORI
end sequence: BH278159"
misc_feature
199680..200563
/notes="clone_boundary"
clone_end:5p6
site:ECORI
end sequence: BH278164"
misc_feature
201821..202949
/notes="wgs_end_extension"
clone_end:5p6"
206162..208510
/notes="wgs_end_extension"
clone_end:5p6"
ORIGIN
Query Match 18.0%; Score 90.2; DB 2; Length 211409;
Best Local Similarity 62.3%; Pred. No. 9.3e-18;
Matches 175; Conservative 0; Mismatches 103; Indels 3; Gaps 2;
QY 221 TTACTTGAATGCTGTGAGTGAAGTGGTCTCTGTCGACAGTAAATCGAGAGGCAATGT 280
DB 155289 TAAATTGAGCGCGGTGGAGAGATGATGCTGTGTCGATTAAGATTAAAGAGCAATAC 155348
QY 281 GCCGTTTGAATGATCTGATGAG-CTACCCCAACAGACAGATGTCAGTCAGGCCCAAT 339
DB 155349 ACCCATTAAGTAAATTAATGAAGGCTTATGTGAACCGGAGGTTTGTCAATAGAGCAGAT 155408
QY 340 CCGTTTGAGTTGATGAGCGGCGGAGACAGTGAAGACACACCTCACTCAGTTTGAGGA 399
DB 155409 CAGATTGAGTTTATGATGGGCAACCAATCAAGAAACAGACACCTCAGCTTGGAAT 155468
QY 400 AGATGAGAACATATGCGCAGAGGATTTGAGAGACCTCGACAGGCGCAAGCCCTAACAC 459
DB 155469 GGAGGATGAAGATACGATTGATG- TTCCGAGCAGACAGACAGGTGTCTACTAATAA 155526
QY 460 ACACACCTACACCTCACTTGAACAAGCTGCAACCAACCA 500
DB 155527 GGGAACTCTACTTACTCCAGAGTTTGTATAGACAA 155567
RESULT 11
AC123449 241990 bp DNA linear HTG 21-SEP-2002
LOCUS Rattus norvegicus clone CH230-21D6, *** SEQUENCING IN PROGRESS ***,
DEFINITION 3 unordered pieces.
ACCESSION AC123449
VERSION AC123449.3 GI:23265406
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 241990)
REFERENCE Murthy D, Marle, M., Meeker, M., Lee, J., Abramson, S., Adams, C., Alder, J.,
AUTHORS Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D.,
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryan, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Creel, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Ditya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Baves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregregis, E., Geer, K., Gill, R., Grady, M., Guertel, W., Guevara, M.,
Gunnarathne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louie, H., Lozano, R.J., Lu, X., Ma, U.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankwelu, O., Okwuonu, G., Olarinmoye, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,
Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
Piazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatk, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wood, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhou, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 241990)
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21902787.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/atlas/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
--- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```

```
----- Project Information
Center project name: GXOF
Center clone name: CH230-21D6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229402 bases at least Q40
Consensus quality: 231763 bases at least Q30
Consensus quality: 233057 bases at least Q20
Estimated insert size: 260790; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_dratc_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 238985: contig of 238985 bp in length
*
238986 239085: gap of unknown length
*
239086 240731: contig of 1646 bp in length
*
240732 240831: gap of unknown length
*
240832 241990: contig of 1159 bp in length.

FEATURES
Source
1..241990
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-21D6"
  1323..2205
  /note="clone boundary
clone end: Sp6
site: EcoRI
end_sequence: BH269859"
misc_feature
  complement(236297..237124)
  /note="clone boundary
clone end: T7
site: EcoRI
end_sequence: BH269858"

ORIGIN
Query Match      18.0%; Score 90.2; DB 2; Length 241990;
Best Local Similarity 62.3%; Pred. No. 9.6e-18;
Matches 175; Conservative 0; Mismatches 103; Indels 3; Gaps 2;

misc_feature
  221 TTAGTTGAATGCTGTGAGTAGTGGGTCTGTGTGGGACAGTAAATCGAGGCGATGT 280
Db      101717 TAAATTGAAGCGGGGAGCAGATGTTCTGTGTGACAGTTAAAGTAAGGCAATAC 101776
Qy      281 GCGGTTTGTGTAAGTGAAG-CTAACCCAAAGACAGAGATTGTCACTGAGGCCAAT 339
Db      101777 ACCACTTACTAATTAATGAAGAGCTAATTTGTAACGCGAGGTTGTCAATGAGCGAGAT 101836
Qy      340 CCGTTTCAGTTGATGGGCGAGCCGACAGTGAAGACACCACTCACTCACTGAGGAGA 399
Db      101837 CAGATTTCAGTTGATGGGCAACCAATCAACGAAACAGACACCTGCAAGTTGGAAT 101896
Qy      400 AGAGTAGAACAATGGCCAGAGAGATTGAGAGACCTGACAGGCGCAAGGCCCTAACAC 459
Db      101897 GAGAGTAGAAGTACGATTGATGTG--TTCCGCGAGAGAGACAGGTGTCTACTAATAAAA 101954
Qy      460 ACACACTTACCACTCTGTAACAAGTGCACAAAGACCAA 500
Db      101955 GGGAACTGCTACTTACTCCAGAGTTTGTATAGAGCAA 101995

RESULT 12
CQ683622      322 bp      DNA      linear      PAT 03-FEB-2004
LOCUS
```

```
DEFINITION Sequence 28548 from Patent WO02070737.
ACCESSION CQ683622
VERSION CQ683622.1 GI:42206548
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Liew,C.C., Marshall,W.E. and Zhang,H.
  Compositions and methods relating to osteoarthritis
  Patent: WO 02070737-A 28548 12-SEP-2002;
  Chondrogene Inc. (CA)
JOURNAL
TITLE
JOURNAL
FEATURES
Source
1..322
  /organism="Homo sapiens"
  /mol_type="unassigned DNA"
  /db_xref="taxon:9606"

ORIGIN
Query Match      17.8%; Score 88.8; DB 6; Length 322;
Best Local Similarity 72.3%; Pred. No. 9.1e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

misc_feature
  221 TTAGTTGAATGCTGTGAGTAGTGGGTCTGTGTGGGACAGTAAATCGAGGCGATGT 280
Db      89 TAAATTGAAGGTGCGGGGAGCAGATGTTCTGTGTGACAGTTTAAAGTAAGGCAATAC 148
Qy      281 GCGGTTTGTGTAAGTGAAG-CTAACCCAAAGACAGAGATTGTCACTGAGGCCAAT 339
Db      149 ACCACTTACTAATTAATGAAGAGCTAATTTGTAACGCGAGGTTGTCAATGAGCGAGAT 208
Qy      340 CCGTTTCAGTTGATGGGCGAGCCGACAGTGAAGACACCACTTCTCACTTGA 356
Db      209 CAGATTCGATTGACGGGCAACCAATCAATGAAGAAACAGACACACTGACAGTTGA 265

RESULT 13
BC083326      933 bp      mRNA      linear      ROD 12-OCT-2004
LOCUS
DEFINITION Mus musculus SMT3 suppressor of mif two 3 homolog 2 (Yeast), mRNA
(CDNA clone MGC:101988 IMAGE:6819700), complete cds.
ACCESSION BC083326
VERSION BC083326.1 GI:54035455
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 933)
  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
  Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
  Altschul,S.F., Zeeberg,B., Buetow,K.H., Scheffer,C.F., Bhat,N.K.,
  Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
  Diatchenko,L., Martinus,K., Farmer,A.A., Rubin,G.M., Hong,L.,
  Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
  Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
  Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
  Abramson,R.D., Mullany,S.J., Bosak,S.A., McMan,P.J.,
  McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richard,S.,
  Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hilyk,S.W.,
  Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
  Fahey,J., Helton,E., Kettelman,M., Madan,A., Young,A.C., Shevchenko,Y.,
  Sanchez,A., Whiting,M., Madan,A., Young,A.C., Rodriques,S.,
  Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
  Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schultz,J., Myers,R.M.,
  Butterfield,Y.S., Krzywinski,M.I., Skalske,V., Smalins,D.E.,
  Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  Generation and initial analysis of more than 15,000 full-length
  human and mouse cDNA sequences
  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL
PUBMED
12477932
2 (bases 1 to 933)
```



**AUTHORS**  
Director MGC Project.  
Direct Submission  
Submitted (01-OCT-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

**REMARK**  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
CDNA Library Preparation: M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [medpaxil.stanford.edu](mailto:medpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>  
Series: IRAK Plate: 195 Row: h Column: 23  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 19111163.  
Location/Qualifiers

**FEATURES**  
source  
1. .933  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="MGC:10198 IMAGE:6819700"  
/tissue\_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"  
/clone\_id="NIH BMAP\_FX0"  
/lab\_host="DH10B"  
/note="Vector: pYX-ASC"  
1. .933  
/gene="Sumo2"  
/note="synonyms: Smt3b, SUMO-2"  
/db\_xref="LocusID:170930"  
/db\_xref="MGI:2158813"  
25. .312  
/gene="Sumo2"  
/codon\_start=1  
/product="SMT3 suppressor of mi f two 3 homolog 2 (yeast)"  
/protein\_id="AAH83326.1"  
/db\_xref="GI:54035456"  
/db\_xref="LocusID:170930"  
/db\_xref="MGI:2158813"  
/translation="MADEKPEKGVKTENNHNINIKVAGDGSVVOFKIKRHTPLSKLM  
KAYCERQGLSMRQIRFRFDGPINETDTPAQLMEDEDTIDVFQOQTGGVY"  
CDS

# ORIGIN

Query Match 17.7%; Score 88.6; DB 10; Length 933;  
Best Local Similarity 61.9%; Pred. No. 1.3e-17;  
Matches 174; Conservative 0; Mismatches 104; Indels 3; Gaps 2;  
QY 221 TTAGTTGAATGATGTGAGTAGAGTGGTCTGTGTGGACAGTAGAATAATCGAGAGCATGT 280  
DB 78 TAAATTGAAGTGGCGGACAGATGTTCTGTGTGTCAGTTTAAAGTTAAGAGCATAC 137  
QY 281 GCCGTTTGAAGCTGATGAG-CTAACCCCAACAGACAGATTTGTCACTCAGTCCCAAT 339  
DB 138 ACCACTTAGTAACTAATGAAGCCTATTGTGAACGACAGGTTTGTCAATGAGGAGAT 197  
QY 340 CCGTTTGAAGTTTATGAGGCGGACAGTAGACAGACACACTCTACTCAGTTGGAGA 399  
DB 198 CAGATTCCGTTTATGAGGCGCAATCAAGAAACAGACACACTCTCAGTTGGAAAT 257  
QY 400 AGGATGAGAACAAATGCGCAGAGGATTTGAGAGACCTTGAAGGCGCAAGCCCTAACAC 459  
DB 258 GGAGGATGAAGATACGATTGATG- -TTCACAGCAGACAGTGGAGGTGTCTACTAAAAA 315  
QY 460 ACACACCTACCACTCACTTGACAAAGCTGCCAAAGACCAA 500

Db

316 GGGAGCTCTCTACTTACTCCAGAAATTTGTATAGACCAA 356

RESULT 14  
MUSMST3A  
LOCUS  
DEFINITION  
MUS musculus MIF2 suppressor (Smt3) mRNA, complete cde.  
VERSION  
L79948.1 GI:17467271  
KEYWORDS  
MIF2 suppressor.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 987)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
Mol. Biol. Evol. 14 (11), 1081-1087 (1997)  
9364765

**FEATURES**  
source  
1. .987  
Location/Qualifiers

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="BALB/c"  
/sub\_species="domesticus"  
/db\_xref="taxon:10090"  
/sex="male"  
/tissue\_type="brain"  
1. .987  
/gene="Smt3"  
1. .97  
/gene="Smt3"  
/note="putative"  
98. .385  
/gene="Smt3"  
/note="putative"  
/codon\_start=1  
/product="MIF2 suppressor"  
/protein\_id="AAI40136.1"  
/db\_xref="GI:17467272"  
/translation="MADEKPEKGVKTENNHNINIKVAGDGSVVOFKIKRHTPLSKLM  
KAYCERQGLSMRQIRFRFDGPINETDTPAQLMEDEDTIDVFQOQTGGVY"  
3' UTR  
polyA\_signal  
/gene="Smt3"  
/note="putative"  
967. .972  
/gene="Smt3"  
/note="putative"

# ORIGIN

Query Match 17.7%; Score 88.6; DB 10; Length 987;  
Best Local Similarity 61.9%; Pred. No. 1.3e-17;  
Matches 174; Conservative 0; Mismatches 104; Indels 3; Gaps 2;  
QY 221 TTAGTTGAATGATGTGAGTAGAGTGGTCTGTGTGGACAGTAGAATAATCGAGAGCATGT 280  
DB 151 TAAATTGAAGTGGCGGACAGATGTTCTGTGTGTCAGTTTAAATTTAAAGGATAC 210  
QY 281 GCCGTTTGAAGCTGATGAG-CTAACCCCAACAGACAGATTTGTCACTCAGTCCCAAT 339  
DB 211 ACCACTTAGTAACTAATGAAGCCTATTGTGAACGACAGGTTTGTCAATGAGGAGAT 270  
QY 340 CCGTTTGAAGTTTATGAGGCGGACAGTAGACAGACACACTCTACTCAGTTGGAGA 399  
DB 271 CAGATTCCGTTTATGAGGCGCAATCAAGAAACAGACACACTCTCAGTTGGAAAT 330  
QY 400 AGGATGAGAACAAATGCGCAGAGGATTTGAGAGACCTTGAAGGCGCAAGCCCTAACAC 459  
DB 331 GGAGGATGAAGATACGATTGATG- -TTCACAGCAGACAGTGGAGGTGTCTACTAAAAA 388

QY 460 ACACACCTACCACTTCACTTGACAAAGCTGCCAAGACCAA 500  
DB 389 GGAAGCCTGCTACTTACTCCAGAAATTTGTTATAGACCAA 429

## RESULT 15

RAT3MT3A 987 bp mRNA linear ROD 10-DEC-2001  
LOCUS Rat3us norvegicus MIF2 suppressor (SMT3) mRNA, complete cds.  
DEFINITION L79949  
ACCESSION L79949.1 GI:17467394  
VERSION MIF2 suppressor.  
KEYWORDS Rat3us norvegicus (Norway rat)  
SOURCE Rat3us norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
Rattus.  
1 (bases 1 to 987)

REFERENCE 1  
AUTHORS Mannen,H., Tsol,S.C., Krushkal,J.S., Li,W.H. and Li,S.S.  
TITLE The cDNA cloning and molecular evolution of reptile and pigeon  
lactate dehydrogenase isozymes  
JOURNAL Mol. Biol. Evol. 14 (11), 1081-1087 (1997)  
MEDLINE 98031373  
PUBMED 9364765

FEATURES  
source Location/Qualifiers  
1..987  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/sex="male"  
/tissue\_type="testis"  
1..987  
/gene="SMT3"  
1..107  
/gene="SMT3"  
/note="putative"  
108..395  
/gene="SMT3"  
/note="putative"  
/codon\_start=1  
/product="MIF2 suppressor"  
/protein\_id="AA140175.1"  
/db\_xref="GI:17467395"  
/translation="MADEKPKBGVTENNNDHINLKVAGODGSVQPKIKRHTPLSKLM  
KAYCERQGLSMRQIRFRFDQPINETDTPAQLEMEDEDTIVFQQGTGVY"  
393..987  
/gene="SMT3"  
/note="putative"  
973..978  
/gene="SMT3"  
/note="putative"

gene  
5'UTR  
CDS

3'UTR

polyA\_signal  
ORIGIN

Query Match 17.7%; Score 88.6; DB 10; Length 987;  
Best Local Similarity 61.9%; Pred. No. 1,3e-17;  
Matches 174; Conservative 0; Mismatches 104; Indels 3; Gaps 2;

QY 221 TTAGTTGAATGCTGTGAGTAGGGTGTCTGTGGACAGTAGAAAATCGAGAGCATGT 280  
DB 161 TAAATTGAAGTGCGCGGACAGAGTGCTGTGTGTCAGATTAAAGATTAAAGAGCATAC 220  
QY 281 GCCGTTAGTGAATCTGATGGAAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCCAAT 339  
DB 221 ACCACTTAGTAACTAATGAAGCTAATTGTGAACGCGAGGTTTGTCAATGAGGAGAT 280  
QY 340 CCGTTTGAGTTTGTATGGCAGCGAGACAGTGAAGACACACCTTACTCAGTTGGAGGA 399  
DB 281 CAGATTCCGTTGATGGGCAACCACTCAAGAAACAGACACACTGCAAGTTGGAAAT 340  
QY 400 AGGATGAGAACAATGGCCAGCAGGATTGAAGACCTGAAGGCCGCAAGCCCTTAACAC 459  
DB 341 GGAGATGAAGATTCAGATTGATGTG--TTCACAGCAGACAGGAGGCGTCTACTGAAAA 398

QY 460 ACACACCTACCACTTCACTTGACAAAGCTGCCAAGACCAA 500  
DB 399 GGAAGCCTGCTACTTACTCCAGAAATTTGTTATAGACCAA 439

Search completed: February 10, 2005, 14:09:22  
Job time : 2412.96 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 ; Search time 340.95 Seconds  
(without alignments)  
8681.254 Million cell updates/sec

Title: US-09-926-375B-7\_COPY\_1\_500  
 Perfect score: 500

Sequence: 1 tcgagagtatcttctgacgac...acaagctgccaaagaccaa 500

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 beqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

```
Database :
1: _Geneseqn_15bec04:*
2: _geneseqn1980s:*
3: _geneseqn1900s:*
4: _geneseqn2000s:*
5: _geneseqn2001as:*
6: _geneseqn2001bs:*
7: _geneseqn2002as:*
8: _geneseqn2002bs:*
9: _geneseqn2003as:*
10: _geneseqn2003bs:*
11: _geneseqn2003ds:*
12: _geneseqn2004as:*
13: _geneseqn2004bs:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	17732	3	AA66A300	AA66A300 Lama2/APP
2	500	100.0	20623	3	AA66A294	AA66A294 Lama2/APP
3	88.8	17.8	557	12	ADQ48843	ADQ48843 Human sc
c	88.2	17.6	300	10	ADDA9509	ADDA9509 Human lu
4	88.2	17.6	350	6	ABK45405	ABK45405 cDNA encod
5	88.2	17.6	490	6	ABQ58908	ABQ58908 Human col
c	88.2	17.6	651	6	ABK54968	ABK54968 Human col
6	88.2	17.6	628	6	ABK44920	ABK44920 cDNA encod
7	88.2	17.6	644	5	AA570174	AA570174 DNA encod
8	88.2	17.6	692	13	ACMA0272	ACMA0272 Tumour-a
9	88.2	17.6	774	10	ADDE4834	ADDE4834 Farnesyl
10	88.2	17.6	774	10	ADDE4834	ADDE4834 Farnesyl
11	88.2	17.6	774	12	ADPE82047	ADPE82047 Human se
12	88.2	17.6	1000	5	ABV25085	ABV25085 Human pro
13	88.2	17.6	1000	12	ADP303040	ADP303040 Human ho
14	88.2	17.6	1000	13	ADSG88538	ADSG88538 Human ho
15	88.2	17.6	1000	12	ADSG88538	ADSG88538 Human ho
16	88.2	17.6	1032	6	ABQ54780	ABQ54780 Human ova
17	88.2	17.6	1039	13	ACMA1069	ACMA1069 Tumour-a
18	88.2	17.6	1385	12	ADQ32327	ADQ32327 Human sc
19	86.6	17.0	426	5	AA581221	AA581221 DNA encod
20	84.8	17.0	444	3	AA616398	AA616398 Human pro

	21	83.4	16.7	516	9	ACH46437	Ach46437 Human inf			
	22	83.4	16.7	516	13	ACN36634	Acn36634 Tumour-a			
	23	82.8	16.6	383	8	ABX41318	Abx41318 Bovine ES			
	24	80.8	16.2	277	4	ABA122368	AbA122368 Probe #12			
C	25	80.8	16.2	277	4	ABA67443	AbA67443 Human toe			
C	26	80.8	16.2	277	4	AA147658	AA147658 Probe #16			
C	27	80.8	16.2	277	4	ABA49552	AbA49552 Human bre			
C	28	80.8	16.2	277	4	ABA34572	AbA34572 Probe #12			
C	29	80.8	16.2	277	4	AAK41616	AAk41616 Human bon			
C	30	80.8	16.2	277	4	AAK15870	AAk15870 Human bra			
C	31	80.8	16.2	277	4	ABS41207	ABs41207 Human liv			
C	32	80.8	16.2	277	5	AAI108052	AAi108052 Probe #80			
C	33	80.8	16.2	277	6	ABS15655	ABs15655 Human gen			
C	34	79	15.8	1859	3	AAFC69094	AAfc69094 Human sec			
C	35	79	15.8	1859	3	AAF15591	AAf15591 Human pro			
C	36	77.4	15.5	315	3	AAZ94114	AAz94114 Haematopo			
C	37	73.4	14.7	508	13	ACN38479	ACn38479 Tumour-a			
C	38	73.4	14.7	1037	12	ADQ83862	Adq83862 Human tu			
C	39	73.4	14.7	1037	12	ADQ87307	Adq87307 Human tu			
C	40	73.4	14.7	1037	13	AOB85021	AOB85021 Human tu			
C	41	73.4	14.7	1037	13	ACN38478	ACn38478 Tumour-a			
C	42	73.4	14.7	1037	13	ACN38564	ACn38564 Tumour-a			
C	43	73.4	14.7	3903	4	AAF28761	AAf28761 Human GAB			
C	44	73.4	14.7	16862	2	AAK58060	AAk58060 Genomi c D			
C	45	73.4	14.7	134292	8	ACA64895	ACa64895 Human GAB			
							ALIGNMENTS			
							RESULT 1			
XX	ID	AAc68300	standard; DNA;	17732 BP.						
XX	AC	AAc68300;								
XX	DT	15-SEP-2003 (revised)								
DT	DT	20-FEB-2001 (first entry)								
XX	DE	Lama2/APPA plasmid coding sequence.								
KW	KM	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;								
XX	OS	environmental pollution; pig; ds.								
XX	OS	Mus musculus.								
XX	OS	Escherichia coli.								
XX	OS	Chimeric.								
PN	PN	WO200064247-A1.								
XX	PD	02-NOV-2000.								
XX	PF	20-APR-2000; 2000MO-CA000430.								
XX	PR	23-APR-1999; 99US-0130508P.								
XX	PA	(UYGU-) UNIV GUELPH.								
XX	PI	Foreberg CW, Golovan S, Phillips JP;								
XX	PI	WPI: 2000-687245/67.								
XX	DR	P-P8DB; AAB36263.								
XX	DR									
XX	PT	Transgenic non-human animal for gastrointestinal tract specific								
XX	PT	expression of a protein, preferably phytase, comprises a nucleic acid								
XX	PT	sequence including a heterologous transgene construct encoding the								
XX	PT	protein.								
XX	PS	Claim 14; Fig 23; 152pp; English.								
XX	PS									
CC	CC	The present invention provides transgenic animals which produce desired								
CC	CC	proteins, in this case pigs which expresses phytase in the salivary								
CC	CC	gland. Low phytase production levels result in phytate in the diet being								

CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 15-SEP-2003 to standardise OS field)  
XX

Sequence 17732 BP; 4119 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;

Query Match 100.0%; Score 500; DB 3; Length 17732;

Best Local Similarity 100.0%; Pred. No. 1,3e-154;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCGAGAGTATCTTTGTGACGTGTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
DB 1 TCGAGAGTATCTTTGTGACGTGTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
QY 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAATAGACGAT 120
DB 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAATAGACGAT 120
QY 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGATGCTGTAAGGGTGGACCTGGCTG 180
DB 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGATGCTGTAAGGGTGGACCTGGCTG 180
QY 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGGATGTTAGTTGAATGGTGGAGT 240
DB 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGGATGTTAGTTGAATGGTGGAGT 240
QY 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
DB 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
QY 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAGGCGCAATCCGTTTGAATTTGAGGCA 360
DB 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAGGCGCAATCCGTTTGAATTTGAGGCA 360
QY 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGAGAAAGATGAAACAATGGCCAGC 420
DB 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGAGAAAGATGAAACAATGGCCAGC 420
QY 421 AGGGATTGAGAGACCTGACAGGCGCAAGGCCCTTAACAACAACCTACCACTCTTG 480
DB 421 AGGGATTGAGAGACCTGACAGGCGCAAGGCCCTTAACAACAACCTACCACTCTTG 480
QY 481 ACAAACTGCCAAAGACCA 500
DB 481 ACAAACTGCCAAAGACCA 500

RESULT 2
AAC68294 standard; DNA; 20623 BP.
XX
AC AAC68294;
XX
DT 15-SEP-2003 (revised)
XX
DT 20-FEB-2001 (first entry)
XX
XX Lama2/APPA plasmid coding sequence.
XX
DE Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX
KM environmental pollution; pig; ds.
XX
XX Mus musculus.
XX
OS Escherichia coli.
XX
OS Chimeric.
XX
XX WO200064247-A1.
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CA000430.
XX
XX 23-APR-1999; 99US-0130508P.
```

XX (UYGU-) UNIV GUELPH.  
PA Forsberg CW, Golovan S, Phillips JP;  
XX  
XX WPI; 2000-687245/67.  
DR P-PSDB; AAB36257.  
XX

PT Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein.  
PS

Claim 56; Fig 5; 152pp; English.

CC The present invention provides transgenic animals which produce desired  
CC proteins, in this case pigs which expresses phytase in the salivary  
CC gland. Low phytase production levels result in phytate in the diet being  
CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 15-SEP-2003 to standardise OS field)  
XX

Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match 100.0%; Score 500; DB 3; Length 20623;

Best Local Similarity 100.0%; Pred. No. 1,4e-154;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCGAGAGTATCTTTGTGACGTGTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
DB 1 TCGAGAGTATCTTTGTGACGTGTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
QY 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAATAGACGAT 120
DB 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAATAGACGAT 120
QY 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGATGCTGTAAGGGTGGACCTGGCTG 180
DB 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGATGCTGTAAGGGTGGACCTGGCTG 180
QY 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGGATGTTAGTTGAATGGTGGAGT 240
DB 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGGATGTTAGTTGAATGGTGGAGT 240
QY 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
DB 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
QY 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAGGCGCAATCCGTTTGAATTTGATGCA 360
DB 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAGGCGCAATCCGTTTGAATTTGATGCA 360
QY 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGAGAAAGATGAAACAATGGCCAGC 420
DB 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGAGAAAGATGAAACAATGGCCAGC 420
QY 421 AGGGATTGAGAGACCTGACAGGCGCAAGGCCCTTAACAACAACCTACCACTCTTG 480
DB 421 AGGGATTGAGAGACCTGACAGGCGCAAGGCCCTTAACAACAACCTACCACTCTTG 480
QY 481 ACAAACTGCCAAAGACCA 500
DB 481 ACAAACTGCCAAAGACCA 500
```

#### RESULT 3

ADQ18843

ID ADQ18843 standard; DNA; 557 BP.

XX ADQ18843;

DT 26-AUG-2004 (first entry)

us-09-926-375b-7\_copy\_1\_500.png

XX	vaccine; T-cell; tumour.
OS	Homo sapiens.
PN	US2003194764-A1.
PD	16-OCT-2003.
PE	04-APR-2002; 2002US-00116712.
XX	
XX	
XX	05-APR-2001; 2001US-0282289P.
PR	05-OCT-2001; 2001US-0327511P.
PA	(CORI-) CORIXA CORP.
XX	
P1	Bangur CS, Switzer A;
DR	WPI; 2003-844452/78.
XX	
PT	New isolated polypeptides and polynucleotides useful for diagnosing,
PT	preventing and treating cancer, particularly lung cancer.
PS	Claim 1; SEQ ID NO 241; 250pp; English.
XX	
CC	The invention relates to an isolated polynucleotide (a) comprising any of
CC	the 666 fully defined nucleotide sequences appearing as ADD99269 -
CC	ADD99935, ADD99936 and ADD99938, complements of (a); sequences of at
CC	least 20 contiguous residues of (a); sequences that hybridise to (a)
CC	under highly stringent conditions; sequences having at least 75 or 90%
CC	identity to (a); or degenerate variants of (a). Also included are an
CC	isolated polypeptide (b) (comprising: sequences encoded by the new
CC	polynucleotide; any of the 4 amino acid sequences fully defined in the
CC	specification; or sequences having at least 70 or 90% identity to the
CC	sequence in (a) or (b)), an expression vector comprising the above
CC	polynucleotide operably linked to an expression control sequence, a host
CC	cell transformed or transfected with the above expression vector, an
CC	isolated antibody, or its antigen-binding fragment, that specifically
CC	binds to the above polypeptide, an oligonucleotide that hybridises to the
CC	above-mentioned nucleotide sequences under highly stringent conditions, a
CC	fusion protein comprising at least one polypeptide cited above, detecting
CC	the presence of a cancer in a patient (comprising: obtaining a biological
CC	sample from the patient; contacting the biological sample with a binding
CC	agent that binds to the polypeptide, or with the oligonucleotide cited
CC	above; detecting in the sample an amount of the polypeptide that binds to
CC	the binding agent, or an amount of a polynucleotide that hybridises to
CC	the oligonucleotide; and comparing the amount of polypeptide, or
CC	polynucleotide that hybridises to the oligonucleotide, to a predetermined
CC	cut-off value and then determining the presence of a cancer in the
CC	patient), a method for stimulating and/or expanding T-cells specific for
CC	a tumour protein (comprising contacting T-cells with the above
CC	polypeptide, polynucleotide or antigen-presenting cells that express the
CC	polynucleotide, under conditions and for a time sufficient to permit the
CC	stimulation and/or expansion of T-cells), an isolated T-cell population
CC	comprising T-cells prepared by the method, a composition comprising a
CC	first component selected from physiological carriers and
CC	immunostimulants, and a second component selected from the above
CC	polypeptide, polynucleotide, antibody, fusion protein, T-cell population
CC	and antigen-presenting cells that express the above polypeptide,
CC	stimulating an immune response in a patient (comprising administering to
CC	the patient the above composition) treating lung cancer in a patient
CC	(comprising administering to the patient the above composition and a
CC	diagnostic kit (comprising: at least one oligonucleotide cited above; or
CC	at least one antibody cited above and a detection reagent, where the
CC	detection reagent comprises a reporter group). The composition and
CC	methods are useful in diagnosing, preventing and treating cancer,
CC	particularly lung cancer. The present sequence is a lung cancer-
CC	associated antigen cDNA of the invention.
SQ	Sequence 300 BP; 73 A; 77 C; 54 G; 96 T; 0 U; 0 Other;
Query Match	17.6%; Score 88.2; DB 10; Length 300;
Best Local Similarity	72.3%; Pred No.1.5e-18;
Matches 128; Conservative	0; Mismatches 46; Indels 1; Gaps 1

QY 221 TTAGTTGAATGCTGTGAGTAGGTTGTTCTGTGTGGACAGTAGAAATCGAGAGCCATGT 280  
DB 274 TAATTTGAAGAGTGCGGGGACAGATGTTCTGTGTGTGACGTTTAAGATTAAGAGCCATAC 215  
QY 261 GCCGTTTGTGTAACATGATGAGAG-CTACCCCAAGACAGAGATGTCAGTCAAGCCCAT 339  
DB 214 ACCACTTGTAACTAAATGAAGAGCCTATTGTGACAGACAGGAGATTGTCAATAGAGCAGAT 155  
QY 340 CCGTTTGCAGTTTGTATGGGACCGGACAGTGAAGACAGACACCTACTCAGTTTGA 396  
DB 154 CAGATTCGATTTTGAACGGGACCAATCAATGAAGAAACAGACACCTGCACAGTTTGA 98

RESULT 5  
ABK45405  
ID ABK45405 standard; cDNA; 350 BP.

AC ABK45405;  
XX  
XX 05-JUN-2002 (first entry)  
XX  
XX CDNA encoding colon tumour protein, SEQ ID No 956.  
XX  
XX Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;  
XX gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200212328-A2.  
XX  
XX 14-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US024218.  
XX  
XX 03-AUG-2000; 2000US-0223283P.  
XX PR 28-MAR-2001; 2001US-0279763P.  
XX PR 29-JUN-2001; 2001US-0302051P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX King GE, Meagher MJ, Xu J, Secrist H;  
XX  
XX WPI; 2002-241739/29.  
XX  
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,  
XX PT for diagnosing, preventing, and treating colon cancer, and as markers for  
XX PT the progression of cancer.  
XX  
XX Claim 1; SEQ ID NO 956; 147bp; English.

XX The invention relates to polynucleotides encoding colon tumour proteins,  
XX CC The polynucleotides and encoded polypeptides are useful in pharmaceutical  
XX CC compositions, such as vaccines, for the diagnosis, prevention, and  
XX CC treatment of colon cancer. Polynucleotide sequences may be used as  
XX CC hybridisation probes or primers, and in the design and preparation of  
XX CC ribozyme molecules for inhibiting expression of tumour polypeptides and  
XX CC proteins in tumour cells. The compositions are useful for stimulating an  
XX CC immune response against cancer, particularly for the immunotherapy of  
XX CC colon cancer, and as markers for the progression of cancer. ABK4450-  
XX CC ABK46237 represent coding sequences of human colon tumour proteins of the  
XX CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence  
XX CC data for this patent did not form part of the printed specification but  
XX CC was supplied by the European Patent Office

XX Sequence 350 BP; 110 A; 69 C; 94 G; 77 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 6; Length 350;  
Best Local Similarity 72.3%; Pred. No. 1.6e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTTGAATGCTGTGAGTAGGTTGTTCTGTGTGGACAGTAGAAATCGAGAGCCATGT 280  
DB 274 TAATTTGAAGAGTGCGGGGACAGATGTTCTGTGTGTGACGTTTAAGATTAAGAGCCATAC 215

DB 77 TAATTTGAAGAGTGCGGGGACAGATGTTCTGTGTGAGTTTAAGATTAAGAGCCATAC 136  
QY 261 GCCGTTTGTGTAACATGATGAGAG-CTACCCCAAGACAGAGATGTCAGTCAAGCCCAT 339  
DB 137 ACCACTTGTAACTAAATGAAGAGCCTATTGTGACAGACAGGAGATTGTCAATAGAGCAGAT 196  
QY 340 CCGTTTGCAGTTTGTATGGGACCGGACAGTGAAGACAGACACCTACTCAGTTTGA 396  
DB 197 CAGATTCGATTTTGAACGGGACCAATCAATGAAGAAACAGACACCTGCACAGTTTGA 253

RESULT 6  
ABO58908/C  
ID ABO58908 standard; cDNA; 490 BP.

XX ABO58908;  
XX  
XX 02-AUG-2002 (first entry)  
XX  
XX Human colon cancer related nucleotide sequence SEQ ID No:2603.  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
XX KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200229086-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US030732.  
XX  
XX 02-OCT-2000; 2000US-0237271P.  
XX  
XX (FARB ) BAYER CORP.  
XX  
XX Burgess C, Aistle JH, Carroll E, Catino TV, Dwiwedi P, Molino GA;  
XX PI Thiglinglam A, Lewis ME;  
XX  
XX WPI; 2002-426115/45.  
XX  
XX New isolated nucleic acid that is differentially expressed in cancer  
XX PT tissues useful for determining the presence of colon cancer in a cell or  
XX PT tissue type, and in antisense therapy.  
XX  
XX Claim 1; Fig 1; 796bp; English.

XX ABO56306 to ABO60787 represent isolated nucleic acids (I) differentially  
XX CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
XX CC encoded by the ABO60776 to ABO60787 nucleic acid sequences. (I) can be  
XX CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
XX CC encoded by (I) is useful for detecting cancer in a patient sample, and  
XX CC for detecting the presence or absence of a polynucleotide encoded by a  
XX CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
XX CC from (I) can be used for determining the presence of a nucleic acid which  
XX CC hybridises to (I), and for determining the phenotype of cells in a sample  
XX CC of cells from a patient. (I) is useful for determining the presence of  
XX CC colon cancer in a cell or tissue type, for determining the presence or  
XX CC state of other type of cancer, in antisense therapy, to generate  
XX CC macroarrays on a solid surface, to identify a chromosome on which the  
XX CC corresponding gene resides, and in tissue profiling, forensics, genetic  
XX CC analysis, mapping and diagnostic applications. (I) can be used to raise  
XX CC antibodies, and to screen for peptide analogues and antagonists

XX Sequence 490 BP; 126 A; 110 C; 109 G; 145 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 6; Length 490;  
Best Local Similarity 72.3%; Pred. No. 1.9e-18;  
Matches 128; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 221 TTAGTTGAATGCTGTGAGTAGGTTGTTCTGTGTGGACAGTAGAAATCGAGAGCCATGT 280  
DB 393 TAATTTGAAGAGTGCGGGGACAGATGTTCTGTGTGACGTTTAAGATTAAGAGCCATAC 334

QY 281 GCCGTTTGAAGTGAAG-CTACCCCAACGACAGAGTTGTCACTACGCGCAAT 339  
DB 333 ACCACTTGAAGTGAAGCTTATGTGTGACGAGGATTTGTCATGAGGCGAT 274  
QY 340 CCGTTTGAAGTGAAGCGGACGACAGTGAAGACACCTACTCAGTTTGA 396  
DB 273 CAGATTCCGATTGACGGGACCAATCAATGAAACAGACACTGACAGTTTGA 217

RESULT 7  
ABK54968  
ID ABK54968 standard; cDNA; 561 BP.  
AC ABK54968;  
XX  
XX 18-JUN-2002 (first entry)  
DT  
DE Human colon cancer-associated cDNA, SEQ ID No 438.  
XX  
XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200212280-A2.  
XX  
XX 14-FEB-2002.  
PD  
XX 30-JUL-2001; 2001WO-US023826.  
PF  
XX 03-AUG-2000; 2000US-0223265P.  
PR 02-OCT-2000; 2000US-0237406P.  
XX 20-MAR-2001; 2001US-0277495P.  
PR 03-JUL-2001; 2001US-0302702P.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Pyle RA, Xu J, Secrist H;  
PI  
XX WPI; 2002-257462/30.  
DR  
XX  
XX Novel polynucleotide encoding colon tumor polypeptides, useful as  
PT vaccines for treating colon cancers.  
XX  
XX Claim 1; Page 275; 425pp; English.  
XX  
XX The invention relates to isolated polynucleotides (I) encoding colon  
CC tumour polypeptides (II). (I) is useful for stimulating an immune  
CC response in a patient and treating colon cancer in a patient.  
CC Oligonucleotides derived from (I) are useful for determining the presence  
CC of cancer in a patient. (I) and (II) are useful in pharmaceutical  
CC compositions, e.g. vaccines, and other compositions for the diagnosis and  
CC treatment of colon cancer. A composition comprising a first component  
CC selected from physiologically acceptable carriers and immunostimulants,  
CC and an antigen-presenting cell expressing (II) is useful for inhibiting  
CC development of cancer in a patient. (I) is useful in the design and  
CC preparation of ribozyme molecules for inhibiting expression of tumour  
CC polypeptides and (I). ABK54531-ABK55464 represent human colon cancer cDNA  
CC sequences of the invention  
XX  
XX Sequence 561 BP; 154 A; 126 C; 134 G; 147 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 6; Length 561;  
Best Local Similarity 72.3%; Pred. No. 2e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTACTTGAATGCTGAGTAGAGTGTCTGTGTGGACAGTAAATCGAGGCGAT 280  
DB 154 TAAATTTGAAGTGGCGGGGAGATGTTCTGTGTGCAATTAAGATTAAGAGGCATAC 213  
QY 281 GCCGTTTGAAGTGAAG-CTACCCCAACGACAGAGATTGTCACTCAGGCGCAAT 339  
DB 214 ACCACTTGAAGTGAAGCTTATGTGTGACGAGGATTTGTCATGAGGCGAT 273

QY 340 CCGTTTGAAGTGAAGCGGACGACAGTGAAGACACACTACTCAGTTTGA 396  
DB 274 CAGATTCCGATTGACGGGACCAATCAATGAAACAGACACTGACAGTTTGA 330

RESULT 8  
ABK44920  
ID ABK44920 standard; cDNA; 628 BP.  
AC ABK44920;  
XX  
XX 05-JUN-2002 (first entry)  
DT  
DE cDNA encoding colon tumour protein, SEQ ID No 471.  
XX  
XX Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;  
XX gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200212328-A2.  
XX  
XX 14-FEB-2002.  
PD  
XX 31-JUL-2001; 2001WO-US024218.  
PF  
XX 03-AUG-2000; 2000US-0223283P.  
PR 28-MAR-2001; 2001US-0279763P.  
XX 29-JUN-2001; 2001US-0302051P.  
PR  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX King GE, Meagher MJ, Xu J, Secrist H;  
PI  
XX WPI; 2002-241739/29.  
DR  
XX  
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,  
PT for diagnosing, preventing, and treating colon cancer, and as markers for  
PT the progression of cancer.  
XX  
XX Claim 1; SEQ ID No 471; 147pp; English.  
XX  
XX The invention relates to polynucleotides encoding colon tumour proteins.  
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical  
CC compositions, such as vaccines, for the diagnosis, prevention, and  
CC treatment of colon cancer. Polynucleotide sequences may be used as  
CC hybridisation probes or primers, and in the design and preparation of  
CC ribozyme molecules for inhibiting expression of tumour polypeptides and  
CC proteins in tumour cells. The compositions are useful for stimulating an  
CC immune response against cancer, particularly for the immunotherapy of  
CC colon cancer, and as markers for the progression of cancer. ABK4450-  
CC ABK46237 represent coding sequences of human colon tumour proteins of the  
CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence  
CC data for this patent did not form part of the printed specification but  
CC was supplied by the European Patent Office  
XX  
XX Sequence 628 BP; 186 A; 132 C; 136 G; 172 T; 0 U; 2 Other;

Query Match 17.6%; Score 88.2; DB 6; Length 628;  
Best Local Similarity 72.3%; Pred. No. 2.1e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTACTTGAATGCTGAGTAGAGTGTCTGTGTGGACAGTAAATCGAGGCGAT 280  
DB 77 TAAATTTGAAGTGGCGGGGAGATGTTCTGTGTGCAATTAAGATTAAGAGGCATAC 136  
QY 281 GCCGTTTGAAGTGAAG-CTACCCCAACGACAGAGATTGTCACTCAGGCGCAAT 339  
DB 137 ACCACTTGAAGTGAAGCTTATGTGTGACGAGGATTTGTCATGAGGCGAT 196  
QY 340 CCGTTTGAAGTGAAGCGGACGACAGTGAAGACACACTACTCAGTTTGA 396

Oy	281	GCCGTTATGCACTGATGGAAG-CTACCCCAACGACAGAGTTGTCAGTCAGGCCAAT	339
Db	224	ACCACTTAGTAACTAAATGAAAGGCTATTGTGAAACGACAGGATTTGTCATGAGGCAAT	283
Oy	340	CCGATTGAGTTGATGGGACGCCGACAGTGAACACACACCTACTCAGTTGGA	396
Db	284	CAGATTCCGATTTCACGGGCAACCAATTAATGAAACAGACACACTCGACAGTTGGA	340
RESULT 10			
ID	ACN40272	standard; cDNA; 692 BP.	
AC	ACN40272;		
DT	18-NOV-2004	(first entry)	
DE	Tumour-associated antigenic target (TAT) cDNA DNA326400, SEQ ID NO:4957.		
XX	Tumour-associated antigenic target; TAT; human; overexpression; cancer;		
XX	tumour; diagnosis; cell proliferative disorder; breast cancer;		
KW	colorectal cancer; lung cancer; ovarian cancer; liver cancer;		
KW	central nervous system cancer; bladder cancer; pancreatic cancer;		
KW	cervical cancer; melanoma; leukemia; hybridisation probe;		
KW	chromosome identification; chromosome mapping; gene mapping;		
KW	gene therapy; cytostatic; gene; ss.		
OS	Homo sapiens.		
XX			
XX	MO2004030615-A2.		
PN	15-APR-2004.		
XX			
XX	29-SEP-2003; 2003WO-US028547.		
PF			
XX	02-OCT-2002; 2002US-0414971P.		
PR			
PA	(GETH ) GENENTECH INC.		
XX			
PI	Wu TD, Zhang Z, Zhou Y;		
XX			
XX	WPI, 2004-347921/32.		
DR			
PT	New tumor-associated antigenic target polypeptides and nucleic acids,		
PT	useful in preparing a medicament for treating or detecting a		
PT	proliferative disorder, e.g. breast, lung, colorectal, ovarian or		
PT	prostate cancer or tumor.		
XX			
XX	Claim 1, SEQ ID NO 4957; 7273BP; English.		
PS			
XX	The invention relates to human tumour-associated antigenic target (TAT)		
CC	polypeptides, and their related nucleic acids. The TAT polypeptides are		
CC	overexpressed in cancer tissues compared to normal tissues, and may thus		
CC	serve as effective targets for the diagnosis and treatment of cancer in		
CC	mammals. The invention also relates to nucleic acid and polypeptide		
CC	sequences at least 80% identical to the TAT nucleic acid and		
CC	polypeptides; expression vectors and host cells comprising a TAT nucleic		
CC	acid; an antibody specific for a TAT polypeptide; fusion proteins comprising a		
CC	molecule which binds to a TAT polypeptide; fusion proteins comprising a		
CC	TAT polypeptide; and methods and compositions for the treatment or		
CC	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,		
CC	antibodies, antagonists, binding molecules and compositions are useful		
CC	for diagnosing or treating a cell proliferative disorder associated with		
CC	increased TAT expression, particularly cancers such as breast cancer,		
CC	colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder		
CC	cancer, pancreatic cancer, cervical cancer, cancers of the central		
CC	nervous system, melanoma and leukaemia. TAT nucleic acids may further be		
CC	used as hybridisation probes, in chromosome and gene mapping, in		
CC	chromosome identification and in gene therapy. The present sequence		
CC	represents a TAT nucleic acid of the invention		
XX			
XX	Sequence 692 BP; 195 A; 143 C; 157 G; 197 T; 0 U; 0 Other;		
SEQ			



Query Match 17.6%; Score 88.2; DB 13; Length 692;  
Best Local Similarity 72.3%; Pred. No. 2.2e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;  
QY 221 TTAGTTGAATGAGTGTGAGTAGGAGTGTCTGTGAGCAGTAGAATAATCGAGGCAATG 280  
DB 134 TAATTTGAAGGTGGCGGGCAGAGATGTTCTGTGTGCGAGTTTAAGATTAAAGGGCATAC 193  
QY 281 GCCGTTTGAATGAATCTGATGAGAG-CTACCCCAAACGACAGAGATTGTCACTCAGGCCAAT 339  
DB 194 ACCACTTAGTAATTAATGAAGCCCTATTGTGAACGACAGGAGATTGTCAATGAGGCAAT 253  
QY 340 CGCTTTCAGTTTGAATGAGGAGCGGACAGCTAGAGACAGACACCTACTAGTTGA 396  
DB 254 CAGATTCCGATTGACGGGACCAATCAATGAACAGACACCTGCAGAGTTGA 310  
RESULT 11  
ADE84834  
ID ADE84834 standard; DNA; 774 BP.  
XX ADE84834;  
AC  
XX  
DT 29-JAN-2004 (first entry)  
DE Farnesyl transferase inhibitor modulated leukemia associated gene #53.  
XX  
XX 88; cytostatic; farnesyl transferase inhibitor; gene expression;  
KM quinolone; leukemia; cancer.  
OS Homo sapiens.  
XX WO2003038129-A2.  
PN  
XX 08-MAY-2003.  
PD  
XX 30-OCT-2002; 2002WO-US034784.  
PF  
XX 30-OCT-2001; 2001US-0338997P.  
PR 30-OCT-2001; 2001US-0340081P.  
PR 30-OCT-2001; 2001US-0340938P.  
PR 30-OCT-2001; 2001US-0341012P.  
XX  
XX (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.  
PA  
PI Rapont M;  
PI WPI; 2003-513497/48.  
DR  
XX Determining whether a patient will respond to treatment with a farnesyl  
PT transferase inhibitor, by analyzing the expression of gene that is  
PT differentially modulated in the presence of the inhibitor.  
PS  
XX Disclosure; SEQ ID NO 53; 346bp; English.  
XX  
XX The invention relates to a method of determining whether a patient will  
CC respond to treatment with a farnesyl transferase inhibitor (FTI), by  
CC analyzing the expression of gene that is differentially modulated in the  
CC presence of an FTI. The method is useful for determining whether a  
CC patient will respond to treatment with a FTI such as (B)-6-(lamino(4-  
CC chlorophenyl)(1-methyl-1H-imidazol-5-yl)methyl)-4-(3-chlorophenyl)-1-  
CC methyl-2-(1H)quinolone, monitoring the therapy of a patient, treating a  
CC patient with leukemia with FTI if the analysis indicates that the patient  
CC will respond. This sequence corresponds to a gene whose expression may be  
CC modulated in the presence of FTI.  
XX  
SQ Sequence 774 BP; 222 A; 166 C; 170 G; 212 T; 0 U; 4 Other;  
Query Match 17.6%; Score 88.2; DB 10; Length 774;  
Best Local Similarity 72.3%; Pred. No. 2.3e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;  
QY 221 TTAGTTGAATGAGTGTGAGTAGGAGTGTCTGTGAGCAGTAGAATAATCGAGGCAATG 280

DB 144 TAATTTGAAGGTGGCGGACAGATGTTCTGTGAGCAGTTTAAGATTAAAGGCAATAC 203  
QY 281 GCCGTTTGAATGAATCTGATGAGAG-CTACCCCAAACGACAGAGATTGTCACTCAGGCCAAT 339  
DB 204 ACCACTTAGTAATTAATGAAGCCCTATTGTGTAACGACAGGAGTTGTCAATGAGGCAAT 263  
QY 340 CGCTTTCAGTTTGAATGAGGAGCGGACAGCTAGAGACAGACACCTACTAGTTGA 396  
DB 264 CAGATTCCGATTGACGGGACCAATCAATGAACAGACACCTGCAGAGTTGA 320  
RESULT 12  
ADP82047  
ID ADP82047 standard; DNA; 774 BP.  
XX ADP82047;  
AC  
XX  
DT 26-AUG-2004 (first entry)  
DE Human sentrin-2 DNA.  
XX  
XX Sentrin-2; SMT3H2; SMT3A; SUMO-3; hyperproliferative disorder; cancer;  
KM cytostatic; gene therapy; human; chromosome 17q25; gene; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 91..378  
FT /tag= a  
FT /product= "Human sentrin-2 protein"  
XX  
XX US2004110151-A1.  
PN  
XX 10-JUN-2004.  
PD  
XX  
PF 10-DEC-2002; 2002US-00316638.  
PR  
XX 10-DEC-2002; 2002US-00316638.  
PR  
XX (ISIS-) ISIS PHARM INC.  
PA  
PI Ward DT, Dobie KW;  
PI WPI; 2004-440340/41.  
DR P-PSDB; ADP82119.  
DR GENBANK; X99585.1.  
DR  
XX  
XX New oligonucleotide compound that inhibits expression of sentrin-2,  
PT useful for preparing a composition for treating hyperproliferative  
PT disorder, e.g. cancer.  
PS  
XX Claim 1; SEQ ID NO 4; 35pp; English.  
XX  
XX The present invention is directed to antisense oligonucleotides targeted  
CC to sentrin-2 (also known as SMT3H2, SMT3A and SUMO-3) and which modulate  
CC the expression of sentrin-2. The invention is useful for preparing a  
CC composition for treating hyperproliferative disorder such as cancer. The  
CC invention acts as a cytostatic agent. The invention is also useful in  
CC gene therapy. The present sequence is human sentrin-2 DNA. The sentrin-2  
CC gene is located on chromosome 17q25.  
XX  
SQ Sequence 774 BP; 222 A; 166 C; 170 G; 212 T; 0 U; 4 Other;  
Query Match 17.6%; Score 88.2; DB 12; Length 774;  
Best Local Similarity 72.3%; Pred. No. 2.3e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;  
QY 221 TTAGTTGAATGAGTGTGAGTAGGAGTGTCTGTGAGCAGTAGAATAATCGAGGCAATG 280  
DB 144 TAATTTGAAGGTGGCGGACAGATGTTCTGTGAGCAGTTTAAGATTAAAGGCAATAC 203  
QY 281 GCCGTTTGAATGAATCTGATGAGAG-CTACCCCAAACGACAGAGATTGTCACTCAGGCCAAT 339



DT 18-NOV-2004 (first entry)  
XX  
DE Human housekeeping gene cDNA sequence SEQ ID NO:81.  
XX  
XX housekeeping gene; human; probe: hybridisation; DNA microarray;  
KM disease investigation; disease diagnosis; gene; ss  
XX  
OS Homo sapiens.  
XX  
PN WO2004035785-A1.  
XX  
PD 29-APR-2004.  
XX  
PF 16-OCT-2002; 2002WO-JP010753.  
XX  
PR 16-OCT-2002; 2002WO-JP010753.  
XX  
PA (NIGA ) NGK INSULATORS LTD.  
XX  
PI Aburatani H, Yamamoto S;  
XX  
DR WPI; 2004-357219/33.  
XX  
PT Human housekeeping and tissue-specific genes, gene sets and transcription  
PT products, probes and microarrays derived from them for reference use in  
PT investigation and diagnosis of disease.  
PS Claim 5; SEQ ID NO 81; 372pp; Japanese.  
XX  
CC The present invention describes 1189 housekeeping genes and gene sets  
CC containing them, expressed in 35 different human tissues. Also described:  
CC (1) RNA and cDNA transcription products of the housekeeping genes; (2)  
CC oligonucleotide probes hybridising to the housekeeping genes; (3) DNA  
CC microarrays containing the oligonucleotide probes; (4) 1704 genes  
CC expressed specifically in particular tissues; (5) RNA and cDNA  
CC transcription products of the tissue-specific genes; (6) oligonucleotide  
CC probes hybridising to the tissue-specific genes; and (7) DNA microarrays  
CC containing the oligonucleotide probes. The tissue-specific genes are  
CC expressed specifically in one of brain, thymus, pituitary, thyroid,  
CC trachea, lung, breast, skin, skeletal muscle, heart, liver, spleen,  
CC kidney, adrenal gland, pancreas, stomach, small intestine, colon,  
CC bladder, prostate gland, testis, ovary, placenta, uterus, bone marrow,  
CC foetal brain or foetal liver. The housekeeping genes can be used as  
CC reference genes for the investigation and diagnosis of disease. The  
CC present sequence represents a human housekeeping gene cDNA sequence which  
CC is given in the exemplification of the present invention.  
XX  
SQ Sequence 1000 BP; 329 A; 184 C; 189 G; 298 T; 0 U; 0 Other;  
Query Match 17.6%; Score 88.2; DB 13; Length 1000;  
Best Local Similarity 72.3%; Pred. No. 2.6e-18;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;  
QY 221 TTAGTTGAATGCTGTGAGTGGTGTCTGTGGACAGTAGAAATCGAGAGCATGT 280  
DB 124 TAAATTTGAAGTGGCGGGCGGAGATGGTCTGTGTGGATTAAGATTAAAGGCAATAC 183  
QY 281 GCCGTTTAGTGAATGATGAG-CTAACCAGACAGAGATTGTCACTCAGGCCAAT 339  
DB 184 ACCACTTAGTAACCTAATGAAGGCTATTGTGAACGACAGGATTTGCAATGAGGCAGAT 243  
QY 340 CCGTTTGCAGTTGATGGCGCAGCCGAGCAGTAGAGACAGACACCTACTCAGTTGA 396  
DB 244 CAGATTGCAATTTGACGGGCAACCAATCAATGAAGACACACCTGCACAGTTGGA 300

Search completed: February 10, 2005, 09:40:59  
Job time : 344.95 secs

THIS PAGE BLANK (USPTO)



```

FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15733
LENGTH: 84587
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(84587)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15733

```

```

Query Match          15.4%; Score 77; DB 4; Length 84587;
Best Local Similarity 68.4%; Pred. No. 9,1e-15;
Matches 121; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

```

```

QY 221 TTAGTTGAATGCTGTGAGTGAAGGCTTCTGTGGACATGAAATCGAGAGCATGT 280
DB 21875 TAATTTGAAGTGGCAGGCGAGGAGGCTTCTGTGTGATTTAGATTAGAGGCATAC 21816
QY 281 GCCGTTAGTGAACCTATGAGAG-CTACCCCAAAGCAGAGATTGTCACTGAGGCCAAT 339
DB 21815 ACCACTTAGTAACTAATGAAACCTGTGTGACATGAGGATGATCAATGAGACAGAT 21756
QY 340 CCGTTTCGAGTTGATGAGCGAGCCGAGCAGTGAAGACACACCTACTCACTTGA 396
DB 21755 CAGATTCGATTGTCAGACGGAACCATCATGAAAGACACACCTGCTCAGTTGA 21639

```

```

RESULT 3
US-09-422-936-60
Sequence 60, Application US/09422936
Patent No. 6465213
GENERAL INFORMATION:
APPLICANT: Ekstrand, Jonas
TITLE OF INVENTION: NEW NUCLEOTIDE SEQUENCES
FILE REFERENCE: 06275-165002
CURRENT APPLICATION NUMBER: US/09/422,936
CURRENT FILING DATE: 1999-10-22
PRIOR APPLICATION NUMBER: US 09/242,608
PRIOR FILING DATE: 1999-02-19
PRIOR APPLICATION NUMBER: PCT/SE98/01947
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: SWEDEN 9703914-2
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: SWEDEN 9800864-2
PRIOR FILING DATE: 1998-03-16
PRIOR APPLICATION NUMBER: SWEDEN 9802575-2
PRIOR FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 60
LENGTH: 15652
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(15652)
OTHER INFORMATION: n = A, T, G, or C
NAME/KEY: misc_feature
LOCATION: (1)...(15652)
OTHER INFORMATION: r = G or A; y = T/U or C; m = A or C;
OTHER INFORMATION: k = G or T/U; s = G or C; w = A or T/U;
OTHER INFORMATION: b = G, C, or T/U; d = A, G, or T/U;

```

```

OTHER INFORMATION: h = A, C, or T/U; v = A, G, or C
NAME/KEY: exon
LOCATION: (3419)...(3444)
NAME/KEY: intron
LOCATION: (3445)...(3908)
NAME/KEY: exon
LOCATION: (3909)...(3993)
NAME/KEY: intron
LOCATION: (3994)...(4694)
NAME/KEY: exon
LOCATION: (4695)...(4898)
NAME/KEY: intron
LOCATION: (4899)...(5652)
NAME/KEY: exon
LOCATION: (5653)...(5838)
NAME/KEY: intron
LOCATION: (5839)...(7184)
NAME/KEY: exon
LOCATION: (7185)...(7205)
NAME/KEY: intron
LOCATION: (7206)...(8310)
NAME/KEY: exon
LOCATION: (8311)...(8806)
NAME/KEY: intron
LOCATION: (8807)...(12271)
NAME/KEY: exon
LOCATION: (12272)...(12406)
NAME/KEY: intron
LOCATION: (12407)...(12820)
NAME/KEY: exon
LOCATION: (12821)...(12991)
NAME/KEY: intron
LOCATION: (12992)...(14089)
NAME/KEY: exon
LOCATION: (14090)...(14191)
NAME/KEY: intron
LOCATION: (14192)...(14477)
NAME/KEY: exon
LOCATION: (14478)...(14543)
NAME/KEY: intron
LOCATION: (14544)...(15002)
NAME/KEY: exon
LOCATION: (15003)...(15194)
NAME/KEY: intron
LOCATION: (15195)...(15652)
US-09-422-936-60

```

```

Query Match          14.7%; Score 73.4; DB 3; Length 15652;
Best Local Similarity 71.2%; Pred. No. 6,4e-14;
Matches 126; Conservative 0; Mismatches 46; Indels 5; Gaps 2;

```

```

QY 221 TTAGTTGAATGCTGTGAGTGAAGGCTTCTGTGGACATGAAATCGAGAGCATGT 280
DB 9 TAATTTGAAGTGGCAGGCGAGGAGTGTGTGTGATTTAGATTAGAGGCATAC 68
QY 281 GCCGTTAGTGAACCTATGAGAG-CTACCCCAAAGCAGAGATTGTCACTGAGGCCAAT 339
DB 69 ACCACTTAGTAACTAATGAAACCTGTGTGACATGAGGATGATCAATGAGACAGAT 128
QY 340 CCGTTTCGAGTTGATGAGCGAGCCGAGCAGTGAAGACACACCTACTCACTTGA 396
DB 129 CAGATTCGATTGTCAGACGGAACCATCATGAAAGACACACCTGCTCAGTTGA 181

```

```

RESULT 4
US-09-949-016-148687/c
Sequence 148687, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307

```

CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 148687  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-148687

Query Match 14.5%; Score 72.4; DB 4; Length 601;  
Best Local Similarity 70.2%; Pred. No. 2.5e-14;  
Matches 125; Conservative 0; Mismatches 51; Indels 2; Gaps 2;

QY 221 TTAGTTGATGTTGAGTGAAGGTTGTTCTGTGGACAGTAAATCGAGAGCCATGT 280  
DB 228 TAAATTGAAGGTGGTGGCGAGGGTGTCTGTGTGAGTTAAGTAAAGAGGCATAC 169  
QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCAAT 339  
DB 168 ACCACTTGAATTAAGTAAAGCCTATTGTGTAATGACAGGATTTGTCATGAGGCAGAT 109  
QY 340 CCGTTTGAAGTTGATGGGCGGCGGA-CAGTGAGACAGACACACTCACTCAGTTGGA 396  
DB 108 CAGACTGTGATTTGACCGGCAACCAATCAATGAAAAAGACACCTGTACAGTTGGA 51

## RESULT 5

US-09-949-016-15921/C  
Sequence 15921, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15921  
LENGTH: 124480  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15921

Query Match 14.5%; Score 72.4; DB 4; Length 124480;  
Best Local Similarity 70.2%; Pred. No. 4.2e-13;  
Matches 125; Conservative 0; Mismatches 51; Indels 2; Gaps 2;

QY 221 TTAGTTGATGTTGAGTGAAGGTTGTTCTGTGGACAGTAAATCGAGAGCCATGT 280  
DB 59936 TAAATTGAAGGTGGTGGCGAGGGTGTCTGTGTGAGTTAAGTAAAGAGGCATAC 59877  
QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCAAT 339  
DB 59876 ACCACTTGAATTAAGTAAAGCCTATTGTGTAATGACAGGATTTGTCATGAGGCAGAT 59817  
QY 340 CCGTTTGAAGTTGATGGGCGGCGGA-CAGTGAGACAGACACACTCACTCAGTTGGA 396  
DB 59816 CAGACTGTGATTTGACCGGCAACCAATCAATGAAAAAGACACCTGTACAGTTGGA 59759

## RESULT 6

US-08-853-974-2  
Sequence 2, Application US/08853974  
Patent No. 5840534  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/853,974  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0289 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 473 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LUNGFET03  
CLONE: 2361410  
US-08-853-974-2

Query Match 14.3%; Score 71.4; DB 2; Length 473;  
Best Local Similarity 67.1%; Pred. No. 4.8e-14;  
Matches 116; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 226 TGAATGTTGAGTGAAGTGAAGTGTCTGTGGACAGTAAATCGAGAGCCATGTCCGT 285  
DB 139 TGAAGTGGCGGCGGAGGAGCGGCTCGTGTGAGTTCAAGATCAAGAGGCACACCGCCG 198  
QY 286 TTAGTGAAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCAATCCGT 344  
DB 199 TGAGCAAGCTGATGAAGAGCCTACTGCGAGAGGAGGAGGCTTGTCAATGAGGCAAGTCAGAT 258  
QY 345 TCGAGTTGATGGGCGGCGGAGCAGTGAAGACACACCTCACTCAGTTGAG 397  
DB 259 TCAGTTTGAAGGCGGCGGAGCCTCAATCAATGAACTCAAGCAGCAGCTGGAG 311

## RESULT 7

US-09-172-988-2  
Sequence 2, Application US/09172988  
Patent No. 6136955  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN  
NUMBER OF SEQUENCES: 3

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/172,988
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/853,974
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0289 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGFET03
CLONE: 2361410
US-09-172-988-2

Query Match      14.3%; Score 71.4; DB 3; Length 473;
Best Local Similarity 67.1%; Pred. No. 4.8e-14;
Matches 116; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY      226 TGAATGTTGATGGAGCGGCTTCTGTGGACATGAGAAATCGAGGCGATGCCGT 285
DB      139 TGAAGTGGCGCGGCGAGACGCTCCGTGGTGCAGTTCAAGATCAAGAGGACACGCCG 198
QY      286 TTAGTAACTGATGAG-CTACCCAAACAGACAGATTGTAGTCAGGCCAATCCGTT 344
DB      199 TGAGCAAGCTGATGAAGGCTTACTGCGAGGCGAGGCGCTTGTCAATGAGGCGATCAGAT 258
QY      345 TCAGTTTGAATGGGCGCGGACGATGAGACAGACACCTACTCAGTTGGAG 397
DB      259 TCAGTTTGAAGGCGGCGACCATGATGAATGAGACTGACACTCCAGCAGCTGGAG 311

RESULT 8
US-09-949-016-5228
Sequence 5228, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5228
```

```

LENGTH: 1728
TYPE: DNA
ORGANISM: Human
US-09-949-016-5228

Query Match      14.3%; Score 71.4; DB 4; Length 1728;
Best Local Similarity 67.1%; Pred. No. 9.6e-14;
Matches 116; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY      226 TGAATGTTGATGGAGCGGCTTCTGTGGACATGAGAAATCGAGGCGATGCCGT 285
DB      146 TGAAGTGGCGCGGCGAGACGCTCCGTGGTGCAGTTCAAGATCAAGAGGACACGCCG 205
QY      286 TTAGTAACTGATGAG-CTACCCAAACAGACAGATTGTAGTCAGGCCAATCCGTT 344
DB      206 TGAGCAAGCTGATGAAGGCTTACTGCGAGGCGAGGCGCTTGTCAATGAGGCGATCAGAT 265
QY      345 TCAGTTTGAATGGGCGCGGACGATGAGACAGACACCTACTCAGTTGGAG 397
DB      266 TCAGTTTGAAGGCGGCGACCATGATGAATGAGACTGACACTCCAGCAGCTGGAG 318

RESULT 9
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9Pt-F15
US-08-232-463-14

Query Match      10.6%; Score 53.2; DB 1; Length 7218;
Best Local Similarity 8.5%; Pred. No. 3.5e-07;
```



Matches 37; Conservative 212; Mismatches 185; Indels 0; Gaps 0;

```
QY 17 CAGCTGCTCCACAAAGGGTACTGTTGCCACATAGAAATCTAAATTAAT 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1522 CATCTATTGAGTTCAAAAACGGCATGTAGCATCTGTAATTAATCATATGCAAG 1463
QY 77 TAATCCTCAGCCGCAATCTTTCAAGCTAAGTTCAGCATGTTGGAAGTTCTC 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1462 TAGTTAAAGATAGAAATTTGTTCCRRRRRRRRRRRRRRRRRRRRRRRR 1403
QY 137 CAAGGAGATACAGATGAGTGGCTATGAGTGCCTGCTGCTAGAGACACCTGC 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1402 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1343
QY 197 ATCTAGTAAGAAAGCAGCGTGTAGTGAATGTGTGAGTGGTGTCTGTGG 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1342 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1283
QY 257 ACAGTAGAAATGAGAGCATGTCCGTTTAGTAAGTATGAACTGACCCCAACGA 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1282 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1223
QY 317 CAGAGATTGTCAGTCAGCGCAATCCGTTTGCATGTGGCAGCCGACAGTAGACA 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1222 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1163
QY 377 GACACACTACTGCTGGAGGAAGATGAGAACATGCGCAGAGGATGAGAGACC 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1162 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1103
QY 437 TGACAGCGCCAGG 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1102 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1089
```

```
RESULT 10
US-09-949-016-15319
; Sequence 15319, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15319
; LENGTH: 18970
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15319
```

```
Query Match 9.1%; Score 45.6; DB 4; Length 18970;
Best Local Similarity 71.4%; Pred. No. 0.00024;
Matches 60; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 221 TTAGTGAATGATGATGAGTGGGTTCTGTGGGACATAGAAATCGAGGACATGT 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3765 TAAATTGAAGGTGGCGGCGAGATGTTCTGTGTGCAATTGAAGTAAAGAGCATAC 3824
QY 281 GCCGTTTAGTGAATGATGATGAGC 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3825 ACCACTTAGTAACTAATGAAAGC 3848
```

RESULT 11

US-08-781-891-79/c

```
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Bn
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620lenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-6031
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-79
```

```
Query Match 9.0%; Score 45.2; DB 3; Length 87350;
Best Local Similarity 66.9%; Pred. No. 0.00072;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
```

```
QY 282 CCGTTAGTGAATGATGAGG-CTACCCCAAGCAGAGATTGTCACTGAGCCCATC 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12102 CCGCTTAGTAACCAATGAAAGCCTATTGTAAAGGGCAGCGTTGCCAATGAGGACATC 12043
QY 341 CGTTTCAGTTTGTATGGGACGCCGACAGTGAAGACACACACTACTCAGTTGGAGG 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12042 AGCTTCAATTGTGACGGCAGCCATCATGAAACAGACACATCTGCACTTTGAAAG 11965
```

RESULT 12

US-09-618-166-79/c

```
; Sequence 79, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Bn
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
```

```

/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.30
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/618,166
/ FILING DATE: 17-Jul-2000
/ CLASSIFICATION: <Unknown>
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: McMaisters, David D.
/ REGISTRATION NUMBER: 33,963
/ REFERENCE/DOCKET NUMBER: 240052.419C1
/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/
/ INFORMATION FOR SEQ ID NO: 79:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 87350 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 79:
/
/ US-09-618-166-79
/
Query Match          9 0%; Score 45.2; DB 4; Length 87350;
Best Local Similarity 66.3%; Pred. No. 0.00072;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 282 CCGTTAGTGAAGTGTGAGG-CTAACCAGGAGATGTGAGTCCCAATC 340
Db 12102 CCGCTTGTAAACCAATGAAAGGCTATGTAAAGGGGACGGTTGCCAATGAGGACAGATC 12043
QY 341 CGTTTCAGTTGATGGGCGCCGGAGACAGTGAACACACACTACTGAGTGGAGG 398
Db 12042 AGGTTCAATTTGACAGGCGAGCCCAATCAATGAACAGACACATCTGCACTTTGGAAG 11985

RESULT 13
US-09-791-211-3/c
/ Sequence 3, Application US/09791211
/ Patent No. 6448080
/ GENERAL INFORMATION:
/ APPLICANT: Donna T. Ward
/ TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
/ FILE REFERENCE: RTS-0205
/ CURRENT APPLICATION NUMBER: US/09/791,211
/ CURRENT FILING DATE: 2001-02-23
/ NUMBER OF SEQ ID NOS: 90
/ SEQ ID NO 3
/ LENGTH: 87543
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 7421
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 7427
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 11609
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 12605
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 12742
```

```

/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 29370
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 29422
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 29979
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 29980
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 29981
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 30136
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 30140
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 31205
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 31206
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 31592
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 33095
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 33160
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 34066
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 34072
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 36816
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 39020
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 42164
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 42459
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 46808
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 46823
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 46826
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 47291
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 52786
/ OTHER INFORMATION: unknown
/ NAME/KEY: unsure
/ LOCATION: 52787
/ OTHER INFORMATION: unknown
```

```

NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68697
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68718
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68733
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68739
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 69785
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79134
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79198
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 86336
OTHER INFORMATION: unknown
OTHER INFORMATION: unknown
US-09-791-211-3

```

```

Query Match          9.0%; Score 45.2; DB 3; Length 87543;
Best Local Similarity 66.9%; Pred. No. 0.00073;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

```

QY 282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGCCATC 340
DB 12295 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGCCATC 12236
QY 341 CGTTGAGTTGATGGGCGGCGGACAGAGACACACTACTGAGTTGGAGG 398
DB 12235 AGTTTCAATTGACAGGCGGCGGACAGAGACACACTACTGAGTTGGAGG 12178

```

```

RESULT 14
US-09-949-001-29/c
Sequence 29, Application US/09949001
Patent No. 682536
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1000789

```

```

CURRENT APPLICATION NUMBER: US/09/949,001
CURRENT FILING DATE: 2003-01-15
PRIOR APPLICATION NUMBER: 60/231,323
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 848
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 143776
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(143776)
OTHER INFORMATION: n = A,T,C or G
US-09-949-001-29

```

```

Query Match          9.0%; Score 45.2; DB 4; Length 143776;
Best Local Similarity 66.9%; Pred. No. 0.00094;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

```

QY 282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGCCATC 340
DB 99899 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGCCATC 99840
QY 341 CGTTGAGTTGATGGGCGGCGGACAGAGACACACTACTGAGTTGGAGG 398
DB 99839 AGTTTCAATTGACAGGCGGCGGACAGAGACACACTACTGAGTTGGAGG 99782

```

```

RESULT 15
US-09-949-001-35/c
Sequence 35, Application US/09949001
Patent No. 682536
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1000789
CURRENT APPLICATION NUMBER: US/09/949,001
CURRENT FILING DATE: 2003-01-15
PRIOR APPLICATION NUMBER: 60/231,323
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 848
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 35
LENGTH: 144034
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(144034)
OTHER INFORMATION: n = A,T,C or G
US-09-949-001-35

```

```

Query Match          9.0%; Score 45.2; DB 4; Length 144034;
Best Local Similarity 66.9%; Pred. No. 0.00094;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

```

QY 282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGCCATC 340
DB 99899 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGCCATC 99840
QY 341 CGTTGAGTTGATGGGCGGCGGACAGAGACACACTACTGAGTTGGAGG 398
DB 99839 AGTTTCAATTGACAGGCGGCGGACAGAGACACACTACTGAGTTGGAGG 99782

```

```

Search completed: February 10, 2005, 17:42:45
Job time : 107.713 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:13:58, Search time 364.334 Seconds  
(Without alignments)  
8097.594 Million cell updates/sec

Title: US-09-926-375B-7\_COPY\_1\_500

Perfect score: 500  
Sequence: 1 tcgagagatccttgcgc.....acaagctcccaagaccac 500

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 537673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.8	17.8	322	17	US-10-242-535A-28548
2	88.8	17.8	322	17	US-10-085-783A-28548
3	88.8	17.8	557	18	US-10-723-860-1662
4	88.2	17.6	300	16	US-10-116-712-241
5	88.2	17.6	303	17	US-10-242-535A-6748
6	88.2	17.6	303	17	US-10-085-783A-6748
7	88.2	17.6	339	17	US-10-057-475B-10573
8	88.2	17.6	339	17	US-10-154-884B-10573
9	88.2	17.6	350	9	US-09-920-300A-956
10	88.2	17.6	350	13	US-10-033-528-956
11	88.2	17.6	350	16	US-10-099-926-956

12	88.2	17.6	461	17	US-10-242-535A-41493	Sequence 41493, A
13	88.2	17.6	461	17	US-10-085-783A-41493	Sequence 41493, A
14	88.2	17.6	465	17	US-10-242-535A-39663	Sequence 39663, A
15	88.2	17.6	465	17	US-10-085-783A-39663	Sequence 39663, A
16	88.2	17.6	490	11	US-09-969-034-2603	Sequence 2603, A
17	88.2	17.6	561	9	US-09-919-580-438	Sequence 438, App
18	88.2	17.6	628	9	US-09-920-300A-471	Sequence 471, App
19	88.2	17.6	628	13	US-10-033-528-471	Sequence 471, App
20	88.2	17.6	628	16	US-10-099-926-471	Sequence 471, App
21	88.2	17.6	774	18	US-10-316-638-4	Sequence 4, App
22	88.2	17.6	774	18	US-10-283-975A-53	Sequence 53, App
23	88.2	17.6	1000	18	US-10-684-422-131	Sequence 131, App
24	88.2	17.6	1000	18	US-10-357-930-25074	Sequence 25074, A
25	88.2	17.6	1032	17	US-10-264-049-660	Sequence 6057, App
26	88.2	17.6	1385	18	US-10-723-860-6057	Sequence 6057, App
27	88.2	17.6	1350	17	US-10-057-475B-10565	Sequence 10565, A
28	85.6	17.1	330	17	US-10-154-884B-10565	Sequence 833, App
29	84.8	17.0	444	9	US-09-925-300-833	Sequence 33649, A
30	83.4	16.7	413	10	US-09-918-995-33649	Sequence 6483, App
31	82.8	16.6	383	9	US-09-960-352-6483	Sequence 19842, A
32	80.8	16.2	277	9	US-09-925-300-26	Sequence 26, App
33	79	15.8	1859	15	US-10-050-882-21	Sequence 21, App
34	78.6	15.7	978	17	US-10-115-831-70	Sequence 70, App
35	77.4	15.5	315	9	US-09-789-919-38	Sequence 38, App
36	73.4	14.7	15652	17	US-10-235-463-60	Sequence 60, App
37	73.4	14.7	134282	17	US-10-240-425-1102	Sequence 1102, App
38	71.2	14.2	490	17	US-10-242-535A-32522	Sequence 32522, A
39	71.2	14.2	490	17	US-10-085-783A-32522	Sequence 13893, A
40	69.8	14.0	413	9	US-09-960-352-13893	Sequence 170, App
41	67.8	13.6	759	9	US-09-910-943-170	Sequence 615, App
42	67.8	13.6	761	9	US-09-910-943-615	Sequence 161, App
43	67.8	13.6	760	9	US-09-910-943-161	Sequence 161, App
44	67.8	13.6	760	9	US-09-910-943-161	Sequence 161, App
45	66.8	13.4	1733	18	US-10-684-422-161	Sequence 161, App

#### ALIGNMENTS

RESULT 1  
US-10-242-535A-28548  
Sequence 28548, Application US/10242535A  
Publication No. US20040013663A1  
GENERAL INFORMATION:  
APPLICANT: ChondroGene Inc.  
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
FILE REFERENCE: 4231/2005  
CURRENT APPLICATION NUMBER: US/10/242,535A  
PRIOR FILING DATE: 2002-09-12  
PRIOR APPLICATION NUMBER: US 10/085,783  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US 60/305,340  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/275,017  
PRIOR FILING DATE: 2001-03-12  
PRIOR APPLICATION NUMBER: US 60/271,955  
NUMBER OF SEQ ID NOS: 58994  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 28548  
LENGTH: 322  
TYPE: DNA  
ORGANISM: Human  
FEATURES:  
NAME/KEY: misc feature  
LOCATION: (21)-(21)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (177)-(177)  
OTHER INFORMATION: n is a, c, g, or t  
US-10-242-535A-28548

```
Query Match      17.8%; Score 88.8; DB 17; Length 322;
Best Local Similarity 72.3%; Pred. No. 3.2e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGAGTGTGAGTAGGGTGTCTGTGGGACAGTAGAAATCGAGAGGCATGT 280
DB      89 TAAATTTGAAGGTGGCGGGCGAGATGTCTGTGGTGCAGTTTAAATTAAGAGGCATAC 148

QY      281 GCCGTTTAACTGATGAGAG-CTACCCCAACGACAGAGATTGTCTAGTCAAGGCCAAT 339
DB      149 ACCACTTAAGTAAACTAATGAAAGCCATATGGGAAACAGACAGGGATGTCTCAATGAGGCAGAT 208

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
DB      209 CAGATTCGATTTGACGGGCAACCAATCAATGAACAGACACCTGCAAGTTGGA 265

RESULT 2
US-10-085-783A-28548
; Sequence 28548, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28548
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (177)..(177)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-28548

Query Match      17.8%; Score 88.8; DB 17; Length 322;
Best Local Similarity 72.3%; Pred. No. 3.2e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGAGTGTGAGTAGGGTGTCTGTGGGACAGTAGAAATCGAGAGGCATGT 280
DB      89 TAAATTTGAAGGTGGCGGGCGAGATGTCTGTGGTGCAGTTTAAATTAAGAGGCATAC 148

QY      281 GCCGTTTAACTGATGAGAG-CTACCCCAACGACAGAGATTGTCTAGTCAAGGCCAAT 339
DB      149 ACCACTTAAGTAAACTAATGAAAGCCATATGGGAAACAGACAGGGATGTCTCAATGAGGCAGAT 208

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
DB      209 CAGATTCGATTTGACGGGCAACCAATCAATGAACAGACACCTGCAAGTTGGA 265

RESULT 3
US-10-723-860-1662
; Sequence 1662, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882, 0193, NUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1662
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (228)..(228)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-1662

Query Match      17.8%; Score 88.8; DB 18; Length 557;
Best Local Similarity 72.3%; Pred. No. 4.1e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGAGTGTGAGTAGGGTGTCTGTGGGACAGTAGAAATCGAGAGGCATGT 280
DB      211 TAAATTTGAAGGTGGCGGGCGAGATGTCTGTGGTGCAGTTTAAATTAAGAGGCATAC 270

QY      281 GCCGTTTAACTGATGAGAG-CTACCCCAACGACAGAGATTGTCTAGTCAAGGCCAAT 339
DB      271 ACCACTTAAGTAAACTAATGAAAGCCATATGGGAAACAGAGAGATTGTCAATGAGGCAGAT 330

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
DB      331 CAGATTCGATTTGACGGGCAACCAATCAATGAACAGACACACCTGCAAGTTGGA 387

RESULT 4
US-10-116-712-241/c
; Sequence 241, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 241
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-116-712-241

Query Match      17.6%; Score 88.2; DB 16; Length 300;
Best Local Similarity 72.3%; Pred. No. 4.9e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGAGTGTGAGTAGGGTGTCTGTGGGACAGTAGAAATCGAGAGGCATGT 280
DB      274 TAAATTTGAAGGTGGCGGGCGAGATGTCTGTGGTGCAGTTTAAATTAAGAGGCATAC 215

QY      281 GCCGTTTAACTGATGAGAG-CTACCCCAACGACAGAGATTGTCTAGTCAAGGCCAAT 339
DB      214 ACCACTTAAGTAAACTAATGAAAGCCATATGGGAAACAGAGAGATTGTCAATGAGGCAGAT 155

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
```



Db	Qy	Db
214	340	154
ACCACTTAGTAAACTATGAAAGCCATTGTGAACAGACAGGGATTGTCAATGAGGCAGAT	CCGTTTCGATGTTGATGGGAGGCGGACAGTGAGACAGACACCTTACTCAGTTGGA	CAGATTCGATTTGACGGGCAACCAATCATATAAACAGACACACTGCACAGTTGGA
155	396	98

RESULT 8  
US-10-15

; Sequence 10573, Application US/10154884B  
; Publication No. US20040005561A1

```

: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: APPLICANT: Mannion, Jane
: APPLICANT: Retter, Marc W.
: APPLICANT: Corixa Corporation
: TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
: TITLE OF INVENTION: Hematological Malignancies
: FILE REFERENCE: 014058-013521US
: CURRENT APPLICATION NUMBER: US/10/154,884B

```

Query Match	17.6%	Score 88.2	DB 17	Length 329
Best Local Similarity	72.3%	Pred. No. 5.2e-19		
Matches 128; Conservative	0	Mismatches 48	Indels 1	Gaps 14

RESULT 9  
US-09-920-300A-956  
; Sequence 956, Application US/09920300A  
; Patent No. US20020136728A1

```

, GENERAL INFORMATION:
, APPLICANT: King, Gordon E
, APPLICANT: Meagher, Madeleine Joy
, APPLICANT: Xu, JiangChun
, APPLICANT: Secrist, Heather
, TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
, TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
, FILE REFERENCE: 210121.547
, CURRENT APPLICATION NUMBER: US/09/920,300A
, CURRENT FILING DATE: 2001-07-31
, NUMBER OF SEQ ID NOS: 1789
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 956
, LENGTH: 350
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-920-300A-956

```

Query Match	17.6%;	Score 88.2;	DB 9;	Length 350;
Best Local Similarity	72.3%;	Pred. No. 5.3e-19;		
Matches 128;	Conservative	0;	Mismatches 48;	Indels 1;
				Gaps 1

```

RESULT 10
US-10-033-528-956
/ Sequence 956, Application US/10033528
/ Publication No. US20020131971A1
/ GENERAL INFORMATION:
/ APPLICANT: King, Gordon E.
/ APPLICANT: Meagher, Madeleine Joy
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Secrist, Heather
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ OF INVENTION: AND DIAGNOSIS OF COLON CANCER
/ FILE REFERENCE: 210121.547C1
/ CURRENT APPLICATION NUMBER: US/10/033,528
/ CURRENT FILING DATE: 2001-12-26
/ NUMBER OF SEQ ID NOS: 1896
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 956
/
/ LENGTH: 350
/
/ TYPE: DNA
/
/ ORGANISM: Homo sapiens
US-10-033-528-956

```

Query Match	17.6%	Score 88.2;	DB 13;	Length 350;
Best Local Similarity	72.3%	Pred. No. 5.3e-19;		
Matches 128; Conservative	0;	Mismatches 48;	Indels 1;	Gaps 1



## RESULT 11

US-10-099-926-956  
; Sequence 956, Application US/10099926  
; Publication No. US2003016064A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Gordon E.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; APPLICANT: Jiang, Yugu  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.547C2  
; CURRENT APPLICATION NUMBER: US/10/099,926  
; CURRENT FILING DATE: 2002-03-17  
; NUMBER OF SEQ ID NOS: 1982  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 956  
; LENGTH: 350  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-099-926-956

Query Match 17.6%; Score 88.2; DB 16; Length 350;  
Best Local Similarity 72.3%; Pred. No. 5.3e-19;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTGAATGAGTGTGAGTGGGTGCTGTGTGGAGACAGTAAATCGAGGCGCATGT 280  
DB 77 TAATTTGAAGGTGGCGGGGAGATGCTGTGTGCTGCACTTTAAGTTAAGGCGATAC 136  
QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTCAGGCCAAT 339  
DB 137 ACCACTTAGTAATTAATGAAGCTTATGTGAACGACAGGAGATTGTCAATGAGGCGAT 196  
QY 340 CCGTTTCGAGTTGAGTGGGACCGGACAGTGAAGACACACCTTCACTGATTGGA 396  
DB 197 CAGATTCCGATTGACCGGCAACCAATCAATGAACAGACACCTGCAAGATTGGA 253

RESULT 12  
US-10-242-535A-41493  
; Sequence 41493, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: IJew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41493  
; LENGTH: 461  
; TYPE: DNA  
; ORGANISM: Human  
US-10-242-535A-41493

Query Match 17.6%; Score 88.2; DB 17; Length 461;  
Best Local Similarity 72.3%; Pred. No. 6e-19;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTGAATGAGTGTGAGTGGGTGCTGTGTGGAGACAGTAAATCGAGGCGCATGT 280

DB 142 TAATTTGAAGGTGGCGGGGAGATGCTGTGTGCTGCACTTTAAGTTAAGGCGATAC 201

QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTCAGGCCAAT 339  
DB 202 ACCACTTAGTAATTAATGAAGCTTATGTGAACGACAGGAGATTGTCAATGAGGCGAT 261

QY 340 CCGTTTCGAGTTGAGTGGGACCGGACAGTGAAGACACACCTTCACTGATTGGA 396  
DB 262 CAGATTCCGATTGACCGGCAACCAATCAATGAACAGACACCTGCAAGATTGGA 318

## RESULT 13

US-10-085-783A-41493  
; Sequence 41493, Application US/10085783A  
; Publication No. US2004003784A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: IJew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41493  
; LENGTH: 461  
; TYPE: DNA  
; ORGANISM: Human  
US-10-085-783A-41493

Query Match 17.6%; Score 88.2; DB 17; Length 461;  
Best Local Similarity 72.3%; Pred. No. 6e-19;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTGAATGAGTGTGAGTGGGTGCTGTGTGGAGACAGTAAATCGAGGCGCATGT 280  
DB 142 TAATTTGAAGGTGGCGGGGAGATGCTGTGTGCTGCACTTTAAGTTAAGGCGATAC 201  
QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTCAGGCCAAT 339  
DB 202 ACCACTTAGTAATTAATGAAGCTTATGTGAACGACAGGAGATTGTCAATGAGGCGAT 261  
QY 340 CCGTTTCGAGTTGAGTGGGACCGGACAGTGAAGACACACCTTCACTGATTGGA 396  
DB 262 CAGATTCCGATTGACCGGCAACCAATCAATGAACAGACACCTGCAAGATTGGA 318

## RESULT 14

US-10-242-535A-39663  
; Sequence 39663, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: IJew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28

Query Match 17.6%; Score 88.2; DB 17; Length 461;  
Best Local Similarity 72.3%; Pred. No. 6e-19;  
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33 ; Search time 2167.44 Seconds  
(without alignments)  
8780.927 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_1\_500

Perfect score: 500

Sequence: 1 tcgagagctctcttgcagc.....acaaagctgcaaaagaccaa 500

Scoring table: IDENTITY NUC

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gest1.\*  
9: gb\_gest2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271.4	54.3	673	8	AZ442128 1M0234N09
2	104.8	21.0	902	2	BF668687 602123353
3	101	20.2	837	2	BF208342 601871604
4	100.2	20.0	696	2	BF208122 601862259
5	98	19.6	797	2	BF675340 602084513
6	98	19.6	853	2	BF698609 602126130
7	97	19.4	710	4	BG614849 602642387
8	97	19.4	837	2	BF028659 601763888
9	97	19.4	866	2	BF958125 601644972
10	96.4	19.3	744	4	BT248345 602966952
11	96	19.2	402	2	BF568319 601341882
12	96	19.2	605	2	BF028876 601765249
13	96	19.2	671	2	BF184845 601843712
14	96	19.2	680	4	BF969520 602271842
15	96	19.2	724	4	BG114143 602285560
16	96	19.2	738	2	BF564680 601343236
17	96	19.2	756	2	BF568240 601341787
18	96	19.2	763	4	BG108259 602280245
19	96	19.2	806	2	BF207530 601862001
20	96	19.2	829	2	BF239978 601905182
21	96	19.2	832	2	BF696841 602125378
22	96	19.2	844	2	BF668313 602122518
23	96	19.2	851	2	BF564620 601342867
24	96	19.2	858	2	BF028632 601763858

25	96	19.2	869	2	BF790681 602250515
26	96	19.2	874	2	BF698481 602123069
27	96	19.2	880	2	BF207800 601862502
28	96	19.2	887	4	BG165717 602344554
29	96	19.2	903	2	BF666162 601679080
30	95.4	19.1	640	4	BM353144 1876b11.x
31	95.4	19.1	663	2	BF212445 601813213
32	95.4	19.1	811	2	BF669919 602118423
33	95.4	19.1	855	2	BF697745 602131295
34	95.4	19.1	892	2	BF695449 601851841
35	95.4	19.1	919	4	BM461727 ACENCOURT
36	94.8	19.0	766	2	BF665640 601677993
37	94.8	19.0	795	4	BG615729 602643106
38	94.8	18.9	825	2	BF698120 602130438
39	94.4	18.9	574	2	BF207801 601862503
40	94.4	18.9	696	2	BF132136 601821166
41	94.4	18.9	788	2	BF568024 601341795
42	94.4	18.9	837	2	BF568054 601341570
43	94	18.8	693	2	BF668853 601678256
44	94	18.8	901	4	BG530611 602560006
45	93.8	18.8	732	2	BF958229 601644694

## ALIGNMENTS

RESULT 1  
AZ442128/c 673 bp DNA linear GSS 03-OCT-2000  
LOCUS 1M0234N09R Mouse 10kb plasmid UGCLIM library Mus musculus genomic  
DEFINITION clone UGCLIM0234N09 R, genomic survey sequence.

ACCESSION AZ442128 GI:10566141  
VERSION AZ442128  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 673)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Rellay,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

JOURNAL  
COMMENT  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
plate: 0234 row: N column: 09  
Seq primer: CACACGAGAAACACTATGACC  
Class: plasmid ends  
High quality sequence stop: 673.  
Location/Qualifiers  
1..673  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCLIM0234N09"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
/clone\_id="Mouse 10kb plasmid UGCLIM library"  
/note="Vector: pWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA

ORIGIN

**QUESTION**

Db 3 3 AGGGTGTTCCTGGGACAGTAGAATA CGAGA 1

mRNA sequence.

ORGANISM    *Homo sapiens*

**COMMENT**      **Contact: Robert Strausberg, Ph.D.**

High quality sequence stop: 597.

```
/mol_type="mRNA"
```

## ORIGIN

Matches 139; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Db 322 CACCTGCACAGTTGA 337

mRNA sequence.

<b>ORGANISM</b>	<b>Homo sapiens</b>
<b>Eukaryotes:</b>	<b>Multicellular:</b>
<b>Chordates:</b>	<b>Cnidarians:</b>
<b>Molluscs:</b>	<b>Invertebrates:</b>
<b>Fungi:</b>	<b>Protists:</b>
<b>Bacteria:</b>	<b>Archaea:</b>
<b>Viruses:</b>	<b>Prions:</b>

**COMMENT**  
**Contact: Robert Strausberg, Ph.D.**

High quality sequence stop: 594.

```
/lab_host="DH10B (T1 phage-resistant)"
```



adaptor sequence: 5'-CAGCGCGCATTAAGGCC-3' and 3' adaptor sequence: 5'-ATTGTGACGCGGCGCGCGCGCATG-dt(30)BN-3' (where B = A, C, G, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match	19.6%	Score 98;	DB 2;	Length 853;
Best Local Similarity	69.2%	Pred. No. 1.4e-18;		
Matches 148; Conservative	0;	Mismatches 65;	Indels 1;	Gaps 1;

QY	201	GACTAGGAAGACCCACGGGTAGTTGTAATAGTGTGAGATGAGGTGGTTCTGTGGACAG	260
Db	146	GACTGGAACACACGATCATTTAATTTTAAAGTGTGGCCGACGAGATGTTCTGTGTGCAG	205
QY	261	TAGAAATCGAAGGCATGTGCGCTTTAGTGAACGTGATGAAG-CTACCCCAACGACAG	319
Db	206	TTTAAATTTAAAGGCATACACCACTTAGTAACTAATGAAGGCTATTGTGAAACGCAG	265
QY	320	AGATTGTCACTAGGCCAATCCGTTTGCAGTTTGATGGCCAGCCGACAATGAGACAGAC	379
Db	266	GGATTGTCAATAGGCAGATCAGATTCGATTTGACGGGCAACCAATCAATGAAACAGAC	325
QY	380	ACACCTACTCAGTTGAGAGAGATGAGAACAT	413
Db	326	ACACCTGCACAGTTGAAATGGAGAGATGAAGAT	359

## RESULT

LOCUS	BC614849	710 bp	mRNA	linear	EST 18-APR-2001
DEFINITION	602642387F1 NIH_MGC_61 Homo sapiens CDNA clone IMAGE:4773332 5', mRNA sequence.				

ACCESSION	EG614849	GI:13666220
VERSION	EG614849.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 710)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLNL at:  
<http://image.llnl.gov>  
Plate: L1CM1644 row: n column: 21  
High quality sequence stop: 640.

FEATURES  
source

```

1..710
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4773332"
/tissue_type="embryonal carcinoma"
/lab_host="H1h10B (T1 phage-resistant)"
/clone_id="NH1 MCC 61"
/notes="Torgan: t66i13; Vector: pDNR-LIB (Clontech); Site_1:
Stil (ggcgctgctggcgc); Site_2: Stil (ggccatcatggcgc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATTGATGCC-3' and 3' adaptor
sequence: 5'-ATTGTATAGCCGACGGCGCCGACATG-dT(30)-BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average

```

Insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 19.4%; Score 97; DB 4; Length 710;  
Best Local Similarity 69.0%; Pred. No. 2.6e-18;  
Matches 147; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY 201 GATTAGGAAGACCGAGCGTGTAGTGAATGCTGAGTGAAGGTGGTCTGTGGGACAG 260  
DB 177 GACTGGAAGACCGAGCGTGTAGTGAATGCTGAGTGAAGGTGGTCTGTGGGACAG 236  
QY 261 TACAAATGAGAGGAGCGTGTAGTGAATGCTGAGTGAAGGTGGTCTGTGGGACAG 319  
DB 237 TTTAAGTTAAGAGCGTGTAGTGAATGCTGAGTGAAGGTGGTCTGTGGGACAG 296  
QY 320 AGATTGTCAGTCAAGCGCAATCCGTTTGAAGTGTGAGTGGGAGCGGACAGTGAACAGAC 379  
DB 297 GGATTGTCAGTCAAGCGGAGTCAAGTTCGATTGAGCGGACCAATCAATGAACAGAC 356  
QY 380 ACACCTACTAGTGTGAGAGAGATGAGAACAA 412  
DB 357 ACACCTGACACGTTTGAAGATGAGATGACAA 389

RESULT 8 837 bp mRNA linear EST 10-OCT-2000  
BF028659  
LOCUS 60176388BF1 NIH\_MGC\_53 Homo sapiens cDNA clone IMAGE:3996165 5',  
DEFINITION mRNA sequence.

ACCESSION BF028659  
VERSION BF028659.1 GI:10736371  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

REFERENCE 1 (bases 1 to 837)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov

Plate: LNCM847 row: P column: 22  
High quality sequence stop: 559.  
Location/Qualifiers

## FEATURES

## source

1. .837  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3996165"  
/ribose\_type="carcinoma, cell line"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_id="NIH\_MGC\_53"  
/note="Organ: bladder; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgctcgcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCGGACATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto, CA)."

## ORIGIN

Query Match 19.4%; Score 97; DB 2; Length 837;  
Best Local Similarity 68.9%; Pred. No. 2.7e-18;  
Matches 133; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 221 TTAAGTGAATGCTGAGTGAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 280  
DB 179 TTAATGAAGTGAAGTGAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 238  
QY 281 GCGGTTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 340  
DB 239 ACCACTTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 298  
QY 341 CGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 400  
DB 299 AGATTCCGATTGAGCGGACCAATCAATGAACAGACACCTGACAGATTGGAATG 358  
QY 401 GGATGGAACAT 413  
DB 359 GAGATGAACAT 371

RESULT 9 866 bp mRNA linear EST 04-OCT-2000  
BE958125  
LOCUS 601644972F1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:3930112 5',  
DEFINITION mRNA sequence.

ACCESSION BE958125  
VERSION BE958125.1 GI:10568830  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

REFERENCE 1 (bases 1 to 866)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov

Plate: LNCM763 row: P column: 17  
High quality sequence stop: 514.  
Location/Qualifiers

## FEATURES

## source

1. .866  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3930112"  
/ribose\_type="primitive neuroectoderm"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_id="NIH\_MGC\_56"  
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgctcgcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCGGACATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

## ORIGIN

Query Match 19.4%; Score 97; DB 2; Length 866;  
Best Local Similarity 71.6%; Pred. No. 2.8e-18;  
Matches 141; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 201 GACTAAGAGAGCCAGCGTGTAGTTGAATGGTGTGAGTACGGTGTCTGTGGACAG 260  
DB 71 GACTGAGAACAGATCATATTAATGAAGTGGCGGGCAGAGATGTTCTGTGTGACAG 130  
QY 261 TGAAGAAATGAGAGGATGTGCGCTTACTGAATCATGATGAAG-CTACCCCAAGACAG 319  
DB 131 TTTAAGATTAAAGGATACACCACTTAGTAACTAATGAAAGCCCTATTGTGACACAG 190  
QY 320 AGATTGTGAGTACGAGCCATCCGTTTCAGTTTGTATGGAGCCGACAGTGAAGACAGAC 379  
DB 191 GGATTGTCAATGAGGAGATCAGATTCGATTGGACGGGACCAATCATTAAGACAGAC 250  
QY 380 ACACCTACTCAGTTGA 396  
DB 251 ACACCTGACAGATTGA 267

RESULT 10  
BI248345 744 bp mRNA linear EST 17-JUL-2001  
LOCUS 60296952F1 NCI\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5116477 5',  
DEFINITION mRNA sequence.  
ACCESSION BI248345  
VERSION BI248345.1 GI:14794150  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS 1 (bases 1 to 744)  
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLNL at:  
http://image.llnl.gov  
Plate: LLM11283 row: h column: 14  
High quality sequence stop: 742.  
Location/Qualifiers  
1..744  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5116477"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Mam2"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

ORIGIN  
Query Match 19.3%; Score 96.4; DB 4; Length 744;  
Best Local Similarity 61.4%; Pred. No. 4e-18;  
Matches 172; Conservative 0; Mismatches 106; Indels 2; Gaps 1;

QY 221 TTAGTTGAATGTTGTGAGTAGGGTGTCTGTGGACAGTAGAAAATCGAGGCAATGT 280  
DB 109 TTAATTGAAGGTGGCGGACAGAGATGTTCTGTGTGACAGTTAAAGATTAAAGGCAATAC 168

QY 281 GCCCTTAGTGAATGATGAGAGCTAACCCCAAGCAGAGATTGTCAATCAGGCCAATC 340  
DB 169 ACCATTGTGTAACATAAGAAAGCCTATGTGAACGGAGGGTTGTCAATGAGGCAATC 228  
QY 341 CTTTTCAGATTGATGGCAGCCGAGCAGTGAAGACACACCACTTCACTCACTTGAAGAA 400  
DB 229 AGATTCCGGTTTGAATGGGACAGCAATCAACAAACAGACACACCTGCACAGTTGGAAATG 288  
QY 401 GGATAGAAACAATGGCCAGCAGGAGTTGAGAGACCCCTGACAGGCGCAAGGCCCTAACCA 460  
DB 289 GAGGATGAAGATACATTGATGTG--TTCCAGACAGCAAGCTGGAAGTGTCTTAATAAAG 346  
QY 461 CACACCTACCACTTCATTGACAAAGCTGCCAAGACCA 500  
DB 347 GGAAGCTGTAATTCTTACTCCAGAAATTTGTTATACCA 386

RESULT 11  
BE568319 402 bp mRNA linear EST 15-AUG-2000  
LOCUS 601341882F1 NIH\_MGC\_53 Homo sapiens cDNA clone IMAGE:3684044 5',  
DEFINITION mRNA sequence.  
ACCESSION BE568319  
VERSION BE568319.1 GI:9812039  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS 1 (bases 1 to 402)  
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Clontech Laboratories, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLNL at: image.llnl.gov  
Plate: LINC370 row: c column: 21  
High quality sequence stop: 402.  
Location/Qualifiers  
1..402  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3684044"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_53"  
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);  
Site 1: SfiI (ggccgctcgagc); Site 2: SfiI  
(ggccatcagcc); Double-stranded cDNA was prepared from  
cell line RNA. 5' and 3' adaptors were used in cloning as  
follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and  
3' adaptor sequence:  
5'-ATTCTAGAGCGGAGAGCGCCGACATG-dT(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.55  
kb (range 0.9-4.0 kb). 15/15 clones contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA)."

ORIGIN  
Query Match 19.2%; Score 96; DB 2; Length 402;  
Best Local Similarity 71.6%; Pred. No. 4.4e-18;  
Matches 126; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 221 TTAGTTGAATGTTGTGAGTAGGGTGTCTGTGGACAGTAGAAAATCGAGGCAATGT 280



Db 159 TAATTGAAGTGGCGGCGAGAGATGTTCTGTGTGTCAGTTTAAGATTAGAGGATAC 218

QY 281 GCCGTTTAGTAACTGATGAGAGCTACCCCAACGACAGAGATTGTGATCAGGCCAATC 340

Db 219 ACCACTTAGTAATGATGAAAGCTATGTGAACGACAGGATGTGCAATGAGGAGATC 278

QY 341 CGTTTGAGTGTGATGGGCGGCGGACAGTGAACAGACACCTACTCTGAGTTTGA 396

Db 279 AGATTCCGATTGACGGGCAACCAATCAATGAACAGACACCTGACAGTTTGA 334

RESULT 12

LOCUS BF028876 605 bp mRNA linear EST 10-OCT-2000

DEFINITION 601765249F1 NIH\_MGC\_53 Homo sapiens cDNA clone IMAGE:397332 5', mRNA sequence.

ACCESSION BF028876

VERSION BF028876.1 GI:10736588

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE 1 (bases 1 to 605)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov](http://image.llnl.gov)

Plate: LNCM851 row: a column: 13

High quality sequence stop: 510.

Location/Qualifiers

1. 605

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:397332"

/tissue\_type="carcinoma, cell line"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_id="NIH\_MGC\_53"

/note="Organ: bladder; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCGGACGATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 19.2%; Score 96; DB 2; Length 605;

Best Local Similarity 71.6%; Pred. No. 5e-18;

Matches 126; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 221 TTAGTTGAATGATGATGAGTGGGATGTTCTGTGGAACAGTAAGAAATCGAGAGCATGT 280

Db 90 TAATTGAAGTGGCGGCGGCGAGATGTTCTGTGTGTCAGTTTAAGATTAGAGGATAC 149

QY 281 GCCGTTTAGTAACTGATGAGAGCTACCCCAACGACAGAGATTGTGATCAGGCCAATC 340

Db 150 ACCACTTAGTAATGATGAAAGCTATGTGAACGACAGGATGTGCAATGAGGAGATC 209

QY 341 CGTTTGAGTGTGATGGGCGGCGGACAGTGAACAGACACCTACTCTGAGTTTGA 396

Db 210 AGATTCCGATTGACGGGCAACCAATCAATGAAAGACAGACACCTGACAGTTTGA 265

RESULT 13

LOCUS BF184845 671 bp mRNA linear EST 31-OCT-2000

DEFINITION 601843712F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4064306 5', mRNA sequence.

ACCESSION BF184845

VERSION BF184845.1 GI:11063266

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE 1 (bases 1 to 671)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LNCM898 row: h column: 03

High quality sequence stop: 612.

Location/Qualifiers

1. 671

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4064306"

/tissue\_type="from chronic myelogenous leukemia"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_id="NIH\_MGC\_54"

/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCGGACGATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 19.2%; Score 96; DB 2; Length 671;

Best Local Similarity 71.6%; Pred. No. 5.2e-18;

Matches 126; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 221 TTAGTTGAATGATGATGAGTGGGATGTTCTGTGGAACAGTAAGAAATCGAGAGCATGT 280

Db 195 TAATTGAAGTGGCGGCGGCGAGATGTTCTGTGTGTCAGTTTAAGATTAGAGGATAC 254

QY 281 GCCGTTTAGTAACTGATGAGAGCTACCCCAACGACAGAGATTGTGATCAGGCCAATC 340

Db 255 ACCACTTAGTAATGATGAAAGCTATGTGAACGACAGGATGTGCAATGAGGAGATC 314

QY 341 CGTTTGAGTGTGATGGGCGGCGGACAGTGAACAGACACCTACTCTGAGTTTGA 396

Db 315 AGATTCCGATTGACGGGCAACCAATCAATGAAAGACAGACACCTGACAGTTTGA 370

RESULT 14

LOCUS BF969520





QY 241 TAAATGTAAGTGGGAGTATGATCTCCAGGTGGGATATCTGCAAGTGCATGAGTGGCA 300  
 DB 5240 TAAATGTAAGTGGGAGTATGATCTCCAGGTGGGATATCTGCAAGTGCATGAGTGGCA 5299  
 QY 301 GAGGAGCAGCAGTATGAGGCAAGAAAGAAATCTGGCTCAACAGCTTAGCTCCCTGGT 360  
 DB 5300 GAGGAGCAGCAGTATGAGGCAAGAAAGAAATCTGGCTCAACAGCTTAGCTCCCTGGT 5359  
 QY 361 GTTGGTCAAACTTTGAGAGTTTGACCAAGACACTTATTTTGTGACATATTTAAACAGA 420  
 DB 5360 GTTGGTCAAACTTTGAGAGTTTGACCAAGACACTTATTTTGTGACATATTTAAACAGA 5419  
 QY 421 GACAACTTTGGGAAAAATTTTCTTATGAAAAATATGCAATTAAGCTTAAGGATGAC 480  
 DB 5420 GACAACTTTGGGAAAAATTTTCTTATGAAAAATATGCAATTAAGGATGAC 5479  
 QY 481 TACATTAATAATGCTTTGCAA 501  
 DB 5480 TACATTAATAATGCTTTGCAA 5500

RESULT 2  
 AX042372 20623 bp DNA linear PAT 23-NOV-2000  
 LOCUS Sequence 1 from Patent WO0064247.  
 DEFINITION AX042372.1 GI:11340990  
 ACCESSION AX042372.1  
 VERSION AX042372.1  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 other sequences; artificial sequences.

REFERENCE  
 1 Forberg, C.W., Golovan, S. and Phillips, J.P.  
 Transgenic animals expressing salivary proteins  
 Patent: WO 0064247-A 1 02-NOV-2000;  
 UNIVERSITY OF GUELPH (CA)

FEATURES  
 source Location/Qualifiers  
 1..20623  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="Lama2/Alpna plasmid"

## ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 20623;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-128;  
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGACCCCATGCTTATTCATGCTTCTTTGCTAGTGTGCTTCTTCTGCTAGTGTGCTT 60  
 DB 5000 GGTGACCCCATGCTTATTCATGCTTCTTTGCTAGTGTGCTTCTTCTGCTAGTGTGCTT 5059  
 QY 61 GTTGGAGCCAGAGAGAGCAGTCCAGGCTGACAGTGTGATACAGAGGAGAGATGATG 120  
 DB 5060 GTTGGAGCCAGAGAGAGCAGTCCAGGCTGACAGTGTGATACAGAGGAGAGATGATG 5119  
 QY 121 GGGTCTCAGCCTGAAGCAGTCAAGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 DB 5120 GGGTCTCAGCCTGAAGCAGTCAAGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 5179  
 QY 181 ACCAATATATCTTTGAAAAAACAATATATATATATATATATATATATATATATATAT 240  
 DB 5180 ACCAATATATCTTTGAAAAAACAATATATATATATATATATATATATATATATATAT 5239  
 QY 241 TAAATGTAAGTGGGAGTATGATCTCCAGGTGGGATATCTGCAAGTGCATGAGTGGCA 300  
 DB 5240 TAAATGTAAGTGGGAGTATGATCTCCAGGTGGGATATCTGCAAGTGCATGAGTGGCA 5299  
 QY 301 GAGGAGCAGCAGTATGAGGCAAGAAAGAAATCTGGCTCAACAGCTTAGCTCCCTGGT 360  
 DB 5300 GAGGAGCAGCAGTATGAGGCAAGAAAGAAATCTGGCTCAACAGCTTAGCTCCCTGGT 5359

QY 361 GTTGGTCAAACTTTGAGAGTTTGACCAAGACACTTATTTTGTGACATATTTAAACAGA 420  
 DB 5360 GTTGGTCAAACTTTGAGAGTTTGACCAAGACACTTATTTTGTGACATATTTAAACAGA 5419  
 QY 421 GACAACTTTGGGAAAAATTTTCTTATGAAAAATATGCAATTAAGCTTAAGGATGAC 480  
 DB 5420 GACAACTTTGGGAAAAATTTTCTTATGAAAAATATGCAATTAAGCTTAAGGATGAC 5479  
 QY 481 TACATTAATAATGCTTTGCAA 501  
 DB 5480 TACATTAATAATGCTTTGCAA 5500

RESULT 3  
 AL732466 172535 bp DNA linear ROD 24-OCT-2002  
 LOCUS Mouse DNA sequence from clone RP23-149G14 on chromosome 2, complete  
 DEFINITION sequence.  
 ACCESSION AL732466  
 VERSION AL732466.7 GI:24394996  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 172535)  
 Sycamore, N.  
 Direct Submission  
 Submitted (20-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Oct 25, 2002 this sequence version replaced gi:21621699.

## COMMENT

Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL, Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-149G14 is from the RPI-23 Mouse PAC library.  
 constructed by the group of Pieter de Jong.  
 For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6.  
 Location/Qualifiers

FEATURES  
 source 1..172535  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-149G14"  
 /clone\_lib="RPI-23"

## ORIGIN

Query Match 100.0%; Score 501; DB 10; Length 172535;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-128;  
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;







gene  
misc\_feature  
/chromosome="2"  
/note="located between -7.7 kb and -6.2 kb in the Mus musculus parotid secretory protein (PSP) gene"  
1..>1557  
/gene="PSP"  
/allele="b"  
/note="synonym: parotid secretory protein gene"  
<1..>1557  
/gene="PSP"  
/allele="b"  
/note="parotid secretory protein"  
/function="salivary gland specific positive acting regulatory region"

## ORIGIN

Query Match 66.5%; Score 333; DB 10; Length 1557;  
Best Local Similarity 100.0%; Pred No. 1.7e-81;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGACCCCATGTGCTTATTCATAGCTTTCTTTGCTATGTTTACTCAGTGTGTTT 60  
DB 1225 GGTGACCCCATGTGCTTATTCATAGCTTTCTTTGCTATGTTTACTCAGTGTGTTT 1284  
QY 61 GTTGGGACCCGAGAGAGCCAGTCCAGGCTGACAGCTGATACAGAGGAGCATGA 120  
DB 1285 GTTGGGACCCGAGAGAGCCAGTCCAGGCTGACAGCTGATACAGAGGAGCATGA 1344  
QY 121 GGGTCTCAGCTGAGAGAGCCAGTCCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 1345 GGGTCTCAGCTGAGAGAGCCAGTCCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1404  
QY 181 ACCAATGATGCTTTGAAAAAACAATATTATTCATATATTTGATTTATGAGACAGC 240  
DB 1405 ACCAATGATGCTTTGAAAAAACAATATTATTCATATATTTGATTTATGAGACAGC 1464  
QY 241 TAAATGTACTGGGTGATGATGATCTCAGGTGGGATATCTGCAATGTCATAGTGGCA 300  
DB 1465 TAAATGTACTGGGTGATGATGATCTCAGGTGGGATATCTGCAATGTCATAGTGGCA 1524  
QY 301 GAGGACAGCCCAATGTGAGGCAAGAGAAATTC 333  
DB 1525 GAGGACAGCCCAATGTGAGGCAAGAGAAATTC 1557

RESULT 7 AC068496 136268 bp DNA linear ROD 15-MAY-2002  
AC068496 LOCUS Mus musculus chromosome 3 clone RP23-75F13 clone RP23-75F13,  
DEFINITION complete sequence.

ACCESSION AC068496 GI:10800151  
VERSION AC068496.7  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 136268)

REFERENCE AUTHORS Montgomey,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
High Throughput Mouse Sequencing

TITLE JOURNAL Unpublished  
REFERENCE JOURNAL 2 (bases 1 to 136268)  
AUTHORS Montgomey,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
Direct Submission  
TITLE Submitted (03-MAY-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA

REFERENCE JOURNAL 3 (bases 1 to 136268)  
AUTHORS Montgomey,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,

## TITLE

Direct Submission  
Submitted (13-SEP-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA

## REFERENCE

4 (bases 1 to 136268)  
Montgomey,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.

## TITLE

Direct Submission  
Submitted (13-OCT-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA

## COMMENT

On Oct 13, 2000 this sequence version replaced gi:9885874.

-----Genome Center:  
Harvard Partners Genome Center  
Center Code: HPGC  
Web site: <http://www.hpcg.org/Sequence/mouse.html>  
Contact: [hpgc@mednet.harvard.edu](mailto:hpgc@mednet.harvard.edu)

CLONE LENGTH: This sequence represents the entire insert of this  
clone unless otherwise noted. If there are overlapping clones, the  
overlaps are noted in the beginning and end of the Features  
listing.

## ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550).  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished) for Human and Mouse sequences.  
Genes and Regions of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST  
and cDNA sequences in Unigene. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintain sequence continuity  
across the splice junctions. Sequences that are not identical  
matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double  
stranded sequence for all regions. All sequence is completed to a  
standard of coverage with a minimum of 3 reads with no ambiguities.  
If the sequence coverage for a region does not meet this standard,  
it is indicated in the annotation as Low Coverage. Low coverage  
linkages are verified by PCR product size verification or  
verification of forward and reverse reads from clones which span  
the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated average error rate is less than 1 per 10,000  
bases using the consed quality parameters. Regions that do not  
meet this requirement are annotated as Low Quality.

## -----Summary Statistics

Center project name: ACM  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100%  
Assembly program: Phrap version 0.990319  
Contig length: 136268  
Fraction of Phrap value < 40: 0.00113012592831773  
Error Rate in Consed: 0.01 per 10,000 bases  
Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

#	Bases
1000	900
900	800
800	700
700	600
600	500
500	400
400	300
300	200
200	100
100	0

\* \* \* \* \*



	repeat_region	24797..24873	/rpt_family="MERCA"
	repeat_region	24895..24943	/rpt_family="(CA)n"
	repeat_region	27749..27882	/complement(27749..27882)
	repeat_region	27905..28002	/rpt_family="B1F"
	repeat_region	28004..28173	/rpt_family="(TA)n"
	repeat_region	28176..28258	/rpt_family="B3"
	repeat_region	29278..29312	/rpt_family="AT_rich"
	repeat_region	29312..29585	/rpt_family="(GA)n"
	repeat_region	30251..30377	/complement(30251..30377)
	repeat_region	31912..32016	/rpt_family="B4A"
	repeat_region	32070..32109	/rpt_family="B1F"
	repeat_region	32109..32137	/rpt_family="(TG)n"
	repeat_region	33092..33331	/complement(33092..33331)
	repeat_region	37560..37752	/rpt_family="B4"
Query Match	Best Local Similarity	26.3%	Score 132; DB 10; Length 136268;
Matches 176;	Conservative 0;	Mismatches 55;	Indels 1; Gaps 1;
OY	270 GTGGGATATTCGACAGTCCATGATGTGGCAGAGGCACCAATGTAGCGCAAGAAGG	329	
Db	118051 GTGGGGGCCCTCAACAATTATGACTGTGAGAGAGAGAGTCTCATGTAGGCAAGAAAG	1179922	
OY	330 ATTTCGGCTCAACACAGCTTAGCTCCCTGTGTGTTCAACTTTGAGAGTTTGACAC	389	
Db	117991 AATCAGTTCAACAGAT-TTAGTTAGCTGTGTGTGGTTCAAACTCTAGAGTGTACC	1179333	
OY	390 AAGCACTTTATTTTGACATATTTAAACAGACACAACTTTGGAAAAAGTTCTTAG	449	
Db	117932 GAGTAGTTATTTTGTGATATCTTAAGCATAGTACAACTTTGAAAAGAGTTCCTTAG	1178733	
OY	450 AAAATTATCACATTAAGCTTAAGCATGACTCACTTAAAGCCCTTGCAA	501	
Db	117872 ACAATTATTAACAATAAGTTTAGACATGACTCACTTAAACTCTTGCAA	117821	
RESULT 8	AC073883/c		
LOCUS	AC073883	213482 bp	DNA linear ROD 14-NOV-2002
DEFINITION	Mus musculus chromosome 3 clone rp23-71e13 strain C57BL/6J,		
VERSION	AC073883		
KEYWORDS	AC073883.34 GI:24961562		
SOURCE	HMG.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus-		
TITLE	1 (bases 1 to 213482)		
JOURNAL	Mus musculus Chromosome 3 BAC Clone rp23-71e13		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 213482)		
TITLE	Hu.X., Swank,R. and Roe,B.A.		
JOURNAL	Direct Submision		
REFERENCE	Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry,		
AUTHORS	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,		
	OK 73019, USA		
	3 (bases 1 to 213482)		
REFERENCE	Hu.X., Swank,R. and Roe,B.A.		

TITLE Direct Submission  
JOURNAL Submitted (12-FEB-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 4 (bases 1 to 213482)  
AUTHORS Hu,X., Swank,R. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-MAR-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 5 (bases 1 to 213482)  
AUTHORS Hu,X., Swank,R. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (14-NOV-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Nov 14, 2002 this sequence version replaced gi:19111728.  
-----  
Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
FEATURES  
source Location/Qualifiers  
1..213482  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/chromosome="3"  
/clone="fp23-71e13"  
/clone\_1ib="RPCT - 23 Female (C57BL/6J) Mouse BAC Library"  
ORIGIN  
Query Match 26.3%; Score 132; DB 10; Length 213482;  
Best Local Similarity 75.9%; Pred. No. 1.2e-25;  
Matches 176; Conservative 0; Mismatches 55; Indels 1; Gaps 1;  
-----  
QY 270 GGGGGATATTCGACATGTCATGTCGACGAGGAGGAGCCATGTGAGGCAAGAAGA 329  
DB 192555 GGGGGGCTCCCTACCAACATTTAGCTGACAGAGATGCTCATGAGGCAAGAA 192496  
QY 330 ATTCTGCTCAACACACAGCTTAGCTCCGCTGGTGGTTCAAACTTTGAGAGTTGACAC 389  
DB 192495 AATCCAGTTCAACAGAT-TTAGTTACTGCTGGTGGTTCAAACTTTGAGAGTTGACAC 192437  
QY 390 AAGCACTTTATTTTGCATATTTAAACAGACAGCAACTTTGGGAAAAAGTTTCTTANG 449  
DB 192436 GAGTAGCTTTATTTTGTATATCTTAAGCATAGTAACTTTGGAAAAAGGTTCTTATG 192377  
QY 450 AAAATTATCATTAAGCTTAAGGACGACATTAATAATGCTTTGCAA 501  
DB 192376 ACAATTATTCATTAAGTTTAAAGAGATGACTAATTAAACCTCTTTGCAA 192325  
-----  
RESULT 9  
AC102009/172592 bp DNA linear ROD 19-AUG-2004  
LOCUS Mus musculus chromosome 10, clone RP24-512D20, complete sequence.  
DEFINITION AC102009  
AC102009.19 GI:51372088  
VERSION HTG.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 172592)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus chromosome 10, clone RP24-512D20  
JOURNAL unpublished  
REFERENCE 2 (bases 1 to 172592)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B.,  
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Grinde,S., Gold,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McNeely,R., Melidre,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., Pollara,V.,  
Oliver,J., Peterson,K., Phunkhupka,P., Pierre,N., Rieck,J.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strause,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.  
Direct Submission  
JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 172592)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
Dearlano,K., Diaz,J.S., Dodge,S., Doolley,K., Dorris,L.,  
Erickson,J., Fato,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Melidre,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhupka,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemdek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
JOURNAL Submitted (08-JUL-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 4 (bases 1 to 172592)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
Dearlano,K., Diaz,J.S., Dodge,S., Doolley,K., Dorris,L.,  
Erickson,J., Fato,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Melidre,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhupka,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemdek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
JOURNAL Submitted (19-AUG-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Aug 19, 2004 this sequence version replaced gi:15058040.

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- genome center  
Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: [mclp://](http://mclp://)

----- Project Information -----

Center project name: L17872

Center clone name: 512\_D\_20

Location/Qualifiers

## FEATURES

[illegible]

```
repeat_region      /rpt_family="L1,F"    complement(27659,.27812)
repeat_region      /rpt_family="L1,F"    complement(27801,.27902)
repeat_region      /rpt_family="L1,F"    complement(27894,.28079)
repeat_region      /rpt_family="L1,n"    complement(28721,.30741)
repeat_region      /rpt_family="L1,M_M"   /rpt_family="L1,M_M"   30746,.30861
repeat_region      /rpt_family="(TA)n"    /rpt_family="(TA)n"    30867,.30899
repeat_region      /rpt_family="(TTTA)n"   complement(30902,.30972)
repeat_region      /rpt_family="ID4"       /rpt_family="ID4"       30990,.31030
repeat_region      /rpt_family="(TCCTA)n"  /rpt_family="(TCCTA)n" 31059,.31127
repeat_region      /rpt_family="(CA)n"     /rpt_family="(CA)n"     31740,.32098
repeat_region      /rpt_family="Lx8"       /rpt_family="Lx8"       32714,.33077
repeat_region      /rpt_family="Lx8"       /rpt_family="Lx8"       33156,.33245
repeat_region      /rpt_family="L1MA8"     /rpt_family="L1MA8"     33407,.33996
unsure             /note=string clone cover
repeat_region      33518,.33653
repeat_region      /rpt_family="GA-rich"    /rpt_family="GA-rich"  33773,.34036
repeat_region      /rpt_family="L1MA8"     /rpt_family="L1MA8"     34037,.34322
repeat_region      /rpt_family="B2_Mm1"    /rpt_family="B2_Mm1"    34228,.34263
repeat_region      /rpt_family="L1MA8"     complement(34281,.34535)
repeat_region      /rpt_family="B4"        /rpt_family="B4"        34536,.34720
```

Query Match	24.6%;	Score 123.4;	DB 10;	Length 172592;
Best Local Similarity	72.4%;	Pred. No. 3e-23;		
Matches 160;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;

QY	281	TGCAAGTCCATGATGCGCAGAGGACACCCAAATGAGGCAGAAAGAAATTCGCTCA	340
Db	32696	TGCAAAATGATACAGAAATGGACAGGAATTAATTCACATGATACAAAAGGGAATTCGTTAA	32637
QY	341	ACACAGCTTAGCTCCCTGGTGTGGTTCAACTTTGAGAGTTTGACACACAGACTTTAT	400
Db	32836	ACACGACTCGGGTTCCTGGTGTGTTGAAATTTCTAGAGCTGAACCTCAACGACTTAT	32577
QY	401	TTTTGACATATTTTAAACAGACCAACCTTTGGAAAAAGTTTTCTTATGAAATATTCAC	460
Db	32576	TTTCAACATTTTAAACACAGTGCACCTTTGGAAAAAAGTTTCTTGACAAATGTTCAC	32517
QY	461	AATTAAGCTTAAGGCATGACTAATTAATAATGCTTTGGCA	501
Db	32516	AAGAAAGCTTAAGGTATGACTACCTTAATCTTTTATCAAA	32476

RESULT 10	AC100513				
LOCUS	AC100513				
DEFINITION	Mus musculus clone RP23-145F3.	70715 bp	DNA	linear	HTG 22-NOV-2001
ACCESSION	AC100513				
VERSION	AC100513.1	GI:17047879			
KEYWORDS	HTG; HTGS PHASE0.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 70715)				
TITLE	Biren,B., Linton,L., Nusbaum,C. and Lander,E.				
	Mus musculus, clone RP23-145F3				

JOURNAL  
REFERENCE  
AUTHORS

Unpublished  
2 (bases 1 to 70715)  
Blitren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barta,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gargay,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
Lamazeas,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPherson,R., Meldrum,J.,  
Menus,L., Minova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunhkhang,P., Piere,N., Pollara,V.,  
Raymond,C., Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Straus,N., Subramanian,A., Talamas,U., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassilleu,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.U., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [submissions@genome.wi.mit.edu](mailto:submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L15511  
Center clone name: 145\_F\_3

-----  
\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
673 772: contig of 672 bp in length  
773 1503: contig of 731 bp in length  
1504 1603: gap of 100 bp  
1604 2348: contig of 745 bp in length  
2349 2448: gap of 100 bp  
2449 3107: contig of 659 bp in length  
3108 3207: gap of 100 bp  
3208 3900: contig of 693 bp in length  
3901 4000: gap of 100 bp  
4001 4704: contig of 704 bp in length  
4705 4804: gap of 100 bp  
4805 5539: contig of 735 bp in length  
5540 5639: gap of 100 bp  
5640 6355: contig of 716 bp in length  
6356 6455: gap of 100 bp  
6456 7198: contig of 743 bp in length  
7199 7298: gap of 100 bp  
7299 8009: contig of 711 bp in length  
8010 8109: gap of 100 bp  
8110 8774: contig of 665 bp in length  
8775 8875 9605: contig of 731 bp in length

\* 9606 9705: gap of 100 bp  
\* 9706 10431: contig of 726 bp in length  
\* 10432 10531: gap of 100 bp  
\* 10532 11261: contig of 730 bp in length  
\* 11262 11361: gap of 100 bp  
\* 11362 12082: contig of 721 bp in length  
\* 12083 12182: gap of 100 bp  
\* 12183 12898: contig of 716 bp in length  
\* 12899 12998: gap of 100 bp  
\* 12999 13709: contig of 711 bp in length  
\* 13710 13809: gap of 100 bp  
\* 13810 14544: contig of 735 bp in length  
\* 14545 14644: gap of 100 bp  
\* 14645 15374: contig of 730 bp in length  
\* 15375 15474: gap of 100 bp  
\* 15475 16161: contig of 742 bp in length  
\* 16217 16316: gap of 100 bp  
\* 16317 17029: contig of 713 bp in length  
\* 17030 17129: gap of 100 bp  
\* 17130 17858: contig of 729 bp in length  
\* 17859 17958: gap of 100 bp  
\* 17959 18623: contig of 665 bp in length  
\* 18624 18723: gap of 100 bp  
\* 18724 19471: contig of 748 bp in length  
\* 19472 19571: gap of 100 bp  
\* 19572 20298: contig of 727 bp in length  
\* 20299 20398: gap of 100 bp  
\* 20399 21142: contig of 744 bp in length  
\* 21143 21242: gap of 100 bp  
\* 21243 21948: contig of 706 bp in length  
\* 21949 22048: gap of 100 bp  
\* 22049 22763: contig of 715 bp in length  
\* 22764 22863: gap of 100 bp  
\* 22864 23585: contig of 722 bp in length  
\* 23586 23685: gap of 100 bp  
\* 23686 24403: contig of 718 bp in length  
\* 24404 25246: contig of 743 bp in length  
\* 25247 25346: gap of 100 bp  
\* 25347 26084: contig of 738 bp in length  
\* 26085 26184: gap of 100 bp  
\* 26185 26926: contig of 742 bp in length  
\* 26927 27026: gap of 100 bp  
\* 27027 27733: contig of 707 bp in length  
\* 27734 27833: gap of 100 bp  
\* 27834 28556: contig of 723 bp in length  
\* 28557 28556: gap of 100 bp  
\* 28557 29393: contig of 737 bp in length  
\* 29394 29493: gap of 100 bp  
\* 29494 30136: contig of 643 bp in length  
\* 30137 30236: gap of 100 bp  
\* 30237 30945: contig of 709 bp in length  
\* 30946 31045: gap of 100 bp  
\* 31046 31768: contig of 723 bp in length  
\* 31769 32688: contig of 820 bp in length  
\* 32689 32788: gap of 100 bp  
\* 32789 33541: contig of 753 bp in length  
\* 33542 33641: gap of 100 bp  
\* 33642 34369: contig of 728 bp in length  
\* 34370 34469: gap of 100 bp  
\* 34470 35193: contig of 724 bp in length  
\* 35194 35293: gap of 100 bp  
\* 35294 36021: contig of 728 bp in length  
\* 36022 36121: gap of 100 bp  
\* 36122 36858: contig of 737 bp in length  
\* 36859 36958: gap of 100 bp  
\* 36959 37679: contig of 721 bp in length  
\* 37680 37779: gap of 100 bp  
\* 37780 38504: contig of 725 bp in length  
\* 38505 38604: gap of 100 bp  
\* 38605 39344: contig of 740 bp in length  
\* 39345 39444: gap of 100 bp

```

* 39445 40175: contig of 731 bp in length
* 40176 40275: gap of 100 bp
* 40276 40935: contig of 660 bp in length
* 40936 41035: gap of 100 bp
* 41036 41749: contig of 714 bp in length
* 41750 41849: gap of 100 bp
* 41850 42582: contig of 733 bp in length
* 42583 42682: gap of 100 bp
* 42683 43418: contig of 736 bp in length
* 43419 44339: contig of 721 bp in length
* 44340 45049: contig of 710 bp in length
* 45050 45149: gap of 100 bp
* 45150 45853: contig of 704 bp in length
* 45854 45953: gap of 100 bp
* 45954 46686: contig of 733 bp in length
* 46687 46787: gap of 100 bp
* 46787 47521: contig of 735 bp in length
* 47522 47621: gap of 100 bp
* 47622 48355: contig of 734 bp in length
* 48356 48455: gap of 100 bp
* 48456 49187: contig of 732 bp in length
* 49188 49287: gap of 100 bp
* 49288 50029: contig of 742 bp in length
* 50030 50129: gap of 100 bp
* 50130 50865: contig of 736 bp in length
* 50866 50965: gap of 100 bp
* 50966 51701: contig of 736 bp in length
* 51702 51801: gap of 100 bp
* 51802 52547: contig of 746 bp in length
* 52548 52647: gap of 100 bp
* 52648 53395: contig of 748 bp in length
* 53396 53495: gap of 100 bp
* 53496 54232: contig of 737 bp in length
* 54233 54332: gap of 100 bp
* 54333 55042: contig of 710 bp in length
* 55043 55142: gap of 100 bp
* 55143 55864: contig of 722 bp in length
* 55865 56705: contig of 741 bp in length
* 56706 56805: gap of 100 bp

```

Query Match 23.7%; Score 118.6; DB 2; Length 70715;  
 Best Local Similarity 69.9%; Pred. No. 6.5e-22;  
 Matches 160; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

QY 256 TAGCATATCTCCAGGTGGGATATCTGCAGTGCATGAGTGGCAGAGGACACGCCATG 315
    |||||
Db 39012 TTGCTTAACGTGATTTTGGGGGTCTCTCCCAACACAGAGTAGGACAGATATGCTGTG 39071
    |||||
QY 316 TGAGGCAAGAAAGAAATCTGGCTCAACAGCCTAGCCCTGGTGGTTCAACTTT 375
    |||||
Db 39072 TGAGGCAAGAAAGCAATCTGTTAAATGTGACTTAAGTAAGTACTGTTCAAACTTC 39131
    |||||
QY 376 GAGAGTTGACCAACAGCACTTATTTTGAATATTTAAACAGAGCAACATTTGGGAA 435
    |||||
Db 39132 TGAGTCTGATGCTGTGAGTATTTTAAACATATTTAAGTATAGCACAACCTTTGGGAA 39191
    |||||
QY 436 AAGCTTTCTTATGAATATTCACATTAAGCTTTAAGGCATGACTTACA 484
    |||||
Db 39192 AGCATTTCTGTGATGATATTCACGATTAAGCTTTAGCCATGACTTACA 39240
    |||||

```

RESULT 11  
 AC135702 185079 bp DNA linear HTG 15-NOV-2002  
 LOCUS Rattus norvegicus clone CH230-419B1, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 2 unordered pieces:  
 AC135702  
 AC135702.2 GI:25007357  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 KEYWORDS Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 185079)  
 Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,  
 Allen,C, Allen,H, Ashbrook,S, Amin,A, Anguiano,D,  
 Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,  
 Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhmed,F,  
 Bielwalo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,  
 Bryant,N, Bunay,C, Burch,P, Burrell,K, Calderon,E,  
 Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,  
 Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,  
 Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,  
 Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,  
 Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,  
 Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,  
 Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,  
 Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,  
 Fraser,C,M, Gabriel,A, Ganta,R, Garcia,A, Garner,T, Garza,M,  
 Gebregeorgis,B, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W,  
 Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,  
 Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,  
 Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogue,M,  
 Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A,  
 Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,  
 Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kover,C,  
 Kowitz,C, Kraft,C,L, Ledow,H, Levan,J, Lewis,L, Li,Z, Liu,J,  
 Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,  
 Lorenshewa,L, Louised,H, Lozada,R,J, Lu,X, Ma,J,  
 Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,  
 Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,  
 Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,  
 Milosavljevic,A, Miner,G, Minja,B, Montemayor,J, Moore,S,  
 Morgan,M, Morris,K, Morris,S, Munidasam,M, Murphy,M, Nair,L,  
 Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,  
 Nwakoelameh,O, Okwuonu,G, Olampunsgoon,A, Pal,S, Parks,K,  
 Pasternak,S, Paul,H, Perez,A, Perez,L, Frankoch,C,  
 Plazzer,F, Polindexter,A, Popovic,D, Primus,E, Pu,L,L,  
 Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,  
 Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,  
 Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,  
 Sanders,W, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,  
 Shetty,D, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajz,D,  
 Sneed,A, Sodegren,E, Song,X,Z, Sorelle,R, Soes,J,  
 Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabors,P, Taylor,C,  
 Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K,  
 Valas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,D,  
 Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,  
 Williams,G, Willson,R, Wlezyk,R, Wooden,H, Worley,K,  
 Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,  
 Yu,F, Zhang,J, Zhou,X, Zhou,X, Zhou,X, Zhou,X, von  
 Niederhausen,A, Weles,R, Smith,D,R, Holt,R,A, Smith,H,O,  
 Weinrock,G, and Gibbs,R,A.

## TITLE

Direct Submission  
 Unpublished  
 2 (bases 1 to 185079)  
 Rat Genome Sequencing Consortium.

REFERENCE  
AUTHORS

Submitted (21-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas

TITLE Direct Submission

JOURNAL Submitted (14-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 249871)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced gi:23673599.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: GUHH  
 Center clone name: CH230-107K7

----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 213554 bases at least Q40  
 Consensus quality: 216552 bases at least Q30  
 Consensus quality: 218611 bases at least Q20  
 Estimated insert size: 217784; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/cenbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/cenbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 247137: contig of 247137 bp in length  
 \* 247138 247237: gap of unknown length  
 \* 247238 248565: contig of 1328 bp in length  
 \* 248566 248665: gap of unknown length  
 \* 248666 249871: contig of 1206 bp in length.

FEATURES  
 source  
 1..249871  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-107K7"

misc\_feature  
 1..1763  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 6697..7589  
 /note="clone\_boundary  
 clone\_end:T7  
 site:  
 end\_sequence:BH270256"

misc\_feature  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 6697..7589  
 /note="clone\_boundary  
 clone\_end:T7  
 site:  
 end\_sequence:BH270256"

ORIGIN  
 Query Match 23.7%; Score 118.6; DB 2; Length 249871;  
 Best Local Similarity 72.7%; Pred. No. 6,4e-22;  
 Matches 168; Conservative 0; Mismatches 59; Indels 4; Gaps 1;

Qy 271 TGGGATATCTGCAAGTGCATGAGGAGGACAGCAATGTGAGGACGAAGAA 330  
 |||||  
 Db 202154 TGGGATCTCCCTGCAATGCGACAGCTGGAGGGTATGATTCGTGAGGACGAAGAAA 202095  
 |||||

Qy 331 TTCTGGCTCAACACAGCTTAGCTCCCTGGTGTGGTTCAAACTTGGAGCTTGACCA 390  
 |||||  
 Db 202094 CTCTGTTACGAAAGATATGACCA---GTGTGATTTCAAACCTTAGAGCTGACCCCA 202039  
 |||||

Qy 391 AGACCTTATTTTGTGACATTTTAAACAGACCAACTTTGGAAAAGTTTCTATGA 450  
 |||||  
 Db 202038 AGCTGTTCCTTTGATATTTTAAACAGACCAACTTTGAAAACAGCTTTCCACGA 201979  
 |||||

Qy 451 AATTATCAATTAAGCTTAAGGACATGACTATTAATAAGCTTTGCA 501  
 |||||  
 Db 201978 CAATGTCAATTAATTAATTAAGGACATGACTATGAAACTCTTACGA 201928  
 |||||

RESULT 13  
 LOCUS AC074315  
 DEFINITION Mus musculus clone rp23-445c23 map 15 strain C57BL/6J, complete sequence.

ACCESSION AC074315  
 VERSION AC074315.77 GI:48331781  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Mus musculus BAC clone rp23-445c23  
 Unpublished  
 2 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (26-JUL-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 3 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (25-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 4 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (25-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 5 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (26-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 6 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (28-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 7 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (03-JUN-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 8 (bases 1 to 177864)  
 Do T., Do A., Hauser, L. and Roe, B.A.  
 Direct Submission  
 Submitted (05-JUN-2004) Department of Chemistry And Biochemistry,

## COMMENT

The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 On Jun 5, 2004 this sequence version replaced gi:48058106.  
 ----- Genome Center  
 Center: Department Of Chemistry And Biochemistry  
 The University of Oklahoma  
 Center code:UDKORH

## FEATURES

source Location/Qualifiers

1. 177864  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="15"  
 /clone="TP23-445c23"  
 /clone\_lib="RPci - 23 Female (C57BL/6J) Mouse BAC library"

## ORIGIN

Query Match 23.5%; Score 117.8; DB 10; Length 177864;  
 Best Local Similarity 68.2%; Pred. No. 1.1e-21;  
 Matches 178; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

241 TAAATGTACTCGGAGATGATGCTCCAGTGGGATATCTGCAAGTCCAGTAGATGCA 300  
 DB TTAAGTACTTATATATCTTTGTCATGCGAGGGTCCCTGTAAGCTTACTGTTGGCA 46675  
 QY 301 GAGGAGACCAATGTGAGGCAAGAAAGATTCTGCTCAACAGCTTACCTCCCTGCT 360  
 DB 46676 GAGAAATTAAACCAAGAGCAAGAAAGAAAGTCTGTCGATGATGACTTGAAGCTAG- 46734  
 QY 361 GTTGTTCAACTTGAAGCTTGAACAGACACTTATTTTTCATATTTAAACACA 420  
 DB 46735 GCTAGTCAAACTCTGAGGTGACATCAAGCAATTTATTTTACATATTTATATCA 46794  
 QY 421 GCACAACCTTGGGAAAAAGTTTCTTATGAAATTTATCAATAAAGCTTAAAGCATGAC 480  
 DB 46795 GTTCAACCTTGAATACAGCTTCTCATGACAAATTCAAATAAAGCTTAAAGCATGAC 46854  
 QY 481 TACATTAAATGCTTTGCA 501  
 DB 46855 TATGTCAAAATTCCTTACAA 46875

RESULT 14  
 AC123698 192333 bp DNA linear HTG 23-SEP-2004  
 LOCUS Mus musculus chromosome 15 clone RP23-331P14 map 15, \*\*\* SEQUENCING  
 DEFINITION IN PROGRESS \*\*\*.

ACCESSION AC123698.11 GI:52546796  
 VERSION HTG; HTGS\_PHASE2; HTGS\_FOLDTOP; HTGS\_ACTIVEPIN.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 192333)  
 AUTHORS Birren, B., Nusbaum, C. and Lander, E.  
 TITLE Mus musculus chromosome 15, clone RP23-331P14  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 192333)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,  
 Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Fero, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K.,  
 Lamatares, R., Landers, T., Lehocsky, J., Levine, R., Lindblad-Ton, K.,  
 Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,  
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,

Meneu, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rie, C.,  
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Saito, R., Schaner, S.,  
 Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,  
 Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## REFERENCE

## AUTHORS

TITLE  
 JOURNAL  
 COMMENT  
 Submitted (23-SEP-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 23, 2004 this sequence version replaced gi:52077830.  
 ALL repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@broad.mit.edu  
 Project Information  
 Center project name: L26429  
 Center clone name: 331\_P\_14

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 192333: contig of 192333 bp in length.

## FEATURES

## source

1. 192333  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="15"  
 /map="15"  
 /clone="RP23-331P14"  
 /clone\_lib="RPci-23 Female Mouse BAC"

## ORIGIN

## Query Match

23.5%; Score 117.8; DB 2; Length 192333;





```

----- Project Information
Center project name: L19783
Center clone name: 315_F_3
-----
FEATURES
source      Location/Qualifiers
            1. 193381
               /organism="Mus musculus"
               /mol_type="genomic DNA"
               /db_xref="taxon:10090"
               /chromosome="15"
               /map="15"
               /clone="RP23-315F3"
               /clone_1ib="RP21-23 Female Mouse BAC"
            1. 8060
               /note="wgs_end_extension"
               /note="wgs_end:SP6"
               /clone_end:SP6"
            701. 736
               /rpt_family="AT_rich"
            748. 1241
               /rpt_family="MER63C"
            1610. 1646
               /rpt_family="AT_rich"
            complement(1735..1802)
               /rpt_family="L1ME1"
            2260. 2394
               /rpt_family="RSINE1"
            2407. 2468
               /rpt_family="CATNA)n"
            2528. 2592
               /rpt_family="TTCC)n"
            complement(3889..4070)
               /rpt_family="B3"
            /rpt_family="Lx8"
            4610. 4692
               /rpt_family="Lx8"
            /rpt_family="Lx8"
            5407. 5485
               /rpt_family="TA)n"
            complement(5522..6201)
               /rpt_family="Lx7"
            8061. 8066
               /note="clone boundary"
               /clone_end:SP6
               site:ECORI"
            8682. 8706
               /rpt_family="TA)n"
            8707. 8743
               /rpt_family="CA)n"
            8735. 8844
               /rpt_family="L1MC4"
            8856. 8876
               /rpt_family="AT_rich"
            complement(8977..9185)
               /rpt_family="URR1A"
            10817. 10851
               /rpt_family="TTTG)n"
            complement(10855..10954)
               /rpt_family="B1_MM"
            complement(11027..11399)
               /rpt_family="ORR1C"
            11617. 11643
               /rpt_family="AT_rich"
            complement(12007..12156)
               /rpt_family="ORR1D"
            14343. 21374
               /rpt_family="L1_MM"
            22887. 22949
               /rpt_family="MER20"
            23208. 23345
               /note="single clone coverage"
            23215. 23220
               /note="<30 qual single clone coverage"
            23238. 23333
               /note="<30 qual single clone coverage"

```

```

/rpt_family="(GAGAA)n"
23275..23290
/note="<30 qual single clone coverage"
23299..23344
/note="<30 qual single clone coverage"
23334..23605
/rpt_family="polypurine"
repeat_region

Query Match      22.9%; Score 114.8; DB 10; Length 193381;
Best Local Similarity 72.1%; Pred. No. 7.3e-21;
Matches 163; Conservative 0; Mismatches 62; Indels 1; Gaps 1;

QY 265 TCCAGGTGGGATATCTGCAAGTGCATGATGTCAGAGGAGCAACCAATGTGAGGCAAG 324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77211 TCATTGGGGGTTTCTCTGCAAAAGCCAGGCTGGCAAGGTAACATCTGATGAGGCTAG 77270
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 325 AAGGAATTCGTGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAAACTTGAAGTTTG 384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77271 AAGGAATTCAGTTCACGTCAGCATAGTTAGCTGTGTGTTCCAACTTGAATTCAG 77330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 385 ACCCAAGCACTTATTTTGGACATATTAAAGAGCAACACTTGGGAAAAAGTTTC 444
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77331 AGCCAAGCACTATATTTTGAACATATTAAACACA-TGCAACTTGGAAAAAGTTTC 77389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 445 TTATGAAATTTATCACAATAAAGCTTAAGGCATGACTTCAATTAAA 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77390 TGGTACAGTTATATATATAAAGCTTAAGGCATGATGACACACAAA 77435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: February 10, 2005, 14:09:29  
Job time : 2420.78 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 ; Search time 341.632 Seconds  
(without alignments)  
8681.254 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_5000\_5500

Perfect score: 501  
Sequence: 1 ggtcagcccatgcttattt.....acattaaatgccttgcac 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseq1960s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002as:\*  
7: geneseq2002bs:\*  
8: geneseq2003as:\*  
9: geneseq2003bs:\*  
10: geneseq2003cs:\*  
11: geneseq2003ds:\*  
12: geneseq2004as:\*  
13: geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	17732	3 AAC68300	Aac68300 Lama2/APP
2	501	100.0	20623	3 AAC68294	Aac68294 Lama2/APP
3	72	14.4	211257	11 ACN44200	Acn44200 Mouse gen
4	69.2	13.8	233060	11 ACN43912	Acn43912 Mouse gen
5	65.6	13.0	110000	8 ABX16390	Abx16390 Mouse can
6	65.2	13.0	110000	8 ABX16390	Abx16390 Mouse hig
7	63.6	12.7	48133	11 ACN44860	Acn44860 Mouse gen
8	61.6	12.3	2893	6 AB199246	Ab199246 Mouse isc
9	61.6	12.3	145068	13 ABD33090	Abd33090 Murine ca
10	61.4	12.3	67832	10 ADA02801	Ada02801 Mouse itk
11	61.4	12.3	67832	10 ABD72539	Abd72539 Mouse itk
12	61.4	12.3	67832	10 ADC85281	Adc85281 Mouse itk
13	61.4	12.3	67832	12 ADM74396	Adm74396 Murine ca
14	60.8	12.1	210528	11 ACN44040	Acn44040 Mouse gen
15	60.4	12.1	121129	13 ABD33446	Abd33446 Murine ca
16	60	12.0	110000	13 ABD32627	Abd32627 Murine ca
17	58.6	11.7	96599	9 ADA02933	Ada02933 Mouse Bra
18	58.6	11.7	96599	10 ABD72671	Abd72671 Mouse Bra
19	58.6	11.7	96599	10 ADC85413	Adc85413 Mouse Bra
20	58.6	11.7	96599	12 ADM74528	Adm74528 Murine ca

## ALIGNMENTS

RESULT 1	
AAC68300	standard; DNA; 17732 BP.
ID	AAC68300
XX	
AC	AAC68300;
XX	
DT	15-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)
XX	
DE	Lama2/APPa plasmid coding sequence.
XX	
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW	environmental pollution; pig; ds.
XX	
OS	Mus musculus.
OS	Escherichia coli.
OS	Chimeric.
XX	
PN	WO200064247-A1.
XX	
PD	02-NOV-2000.
XX	
PF	20-APR-2000; 2000WO-CA000430.
XX	
PR	23-APR-1999; 99US-0130508P.
XX	
PA	(UYGU-) UNIV GUELPH.
XX	
PI	Forsberg CW, Golovan S, Phillips JP;
XX	
DR	WPI; 2000-687245/67.
DR	P-PSDB; AAB36263.
XX	
PT	Transgenic non-human animal for gastrointestinal tract specific
PT	expression of a protein, preferably phytase, comprises a nucleic acid
PT	sequence including a heterologous transgene construct encoding the
PT	protein.
XX	
PS	Claim 14; Fig 23; 152p; English.
XX	
CC	The present invention provides transgenic animals which produce desired
CC	proteins, in this case pigs which express phytase in the salivary
CC	gland. Low phytase production levels result in phytate in the diet being

CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 15-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;

Query Match 100.0%; Score 501; DB 3; Length 17732;

Best Local Similarity 100.0%; Pred. No. 4.6e-138; Indels 0; Gaps 0;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GGTGAGCCCATGCTTATTCATAGCTTCTTTCTTTGTTTCTAGTGTGTTT 60
DB GGTGAGCCCATGCTTATTCATAGCTTCTTTCTTTGTTTCTAGTGTGTTT 5059
QY 61 GTTGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB GTTGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY 121 GGGTCTCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB GGGTCTCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5179
QY 181 ACCAATATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB ACCAATATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5239
QY 241 TAAATATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB TAAATATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5299
QY 301 GAGGAGACAGCAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB GAGGAGACAGCAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5359
QY 361 GTTGGTCAAACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB GTTGGTCAAACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5419
QY 421 GCACAACTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB GCACAACTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5479
QY 481 TACATTAATGCTTTGCA 501
DB TACATTAATGCTTTGCA 5500
```

## RESULT 2

AC68294 standard; DNA; 20623 BP.

```
XX AAC68294;
AC 15-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
XX Lama2/APPA plasmid coding sequence.
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX environmental pollution; pig; ds.
XX
XX Mus musculus.
XX Escherichia coli.
XX Chimeric.
XX
XX WO200064247-A1.
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CAD00430.
XX
XX 23-APR-1999; 99US-0130508P.
XX
```

XX (UYGU-) UNIV GUELPH.  
XX  
XX Forsberg CW, Golojan S, Phillips JP;  
XX  
XX WPI, 2000-687245/67.  
XX  
XX P-PSDB; AAB36257.

PT Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein.  
XX  
PS Claim 56; Fig 5; 152pp; English.

CC The present invention provides transgenic animals which produce desired  
CC proteins, in this case pigs which expresses phytase in the salivary  
CC gland. Low phytase production levels result in phytate in the diet being  
CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 15-SEP-2003 to standardise OS field)  
XX

SQ Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match 100.0%; Score 501; DB 3; Length 20623;

Best Local Similarity 100.0%; Pred. No. 4.9e-138; Indels 0; Gaps 0;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GGTGAGCCCATGCTTATTCATAGCTTCTTTCTTTGTTTCTAGTGTGTTT 60
DB GGTGAGCCCATGCTTATTCATAGCTTCTTTCTTTGTTTCTAGTGTGTTT 5059
QY 61 GTTGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB GTTGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY 121 GGGTCTCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB GGGTCTCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5179
QY 181 ACCAATATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB ACCAATATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5239
QY 241 TAAATATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB TAAATATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5299
QY 301 GAGGAGACAGCAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB GAGGAGACAGCAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5359
QY 361 GTTGGTCAAACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB GTTGGTCAAACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5419
QY 421 GCACAACTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB GCACAACTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5479
QY 481 TACATTAATGCTTTGCA 501
DB TACATTAATGCTTTGCA 5500
```

## RESULT 3

ACN44200/c standard; DNA; 211257 BP.

```
XX ACN44200;
XX
XX 18-NOV-2004 (first entry)
XX
```

XX Mouse genomic sequence mCG15870.  
XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
XX Mus musculus.  
XX WO2003073826-A2.  
XX 12-SEP-2003.  
XX 28-FEB-2003; 2003WO-US006235.  
XX 01-MAR-2002; 2002US-00087192.  
XX (SAGR-) SAGRES DISCOVERY.  
XX Morris DW;  
XX MPI; 2003-328604/31.  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX Claim 1; SEQ ID NO 529; Opp; English.  
XX The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (ix) as a biochip;  
XX carcinoma; (viii) for neutralizing the effect of CAP; and (xi) for  
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published  
XX  
SQ Sequence 211257 BP; 55148 A; 47354 C; 46616 G; 57484 T; 0 U; 4655 Other;  
Query Match 14.4%; Score 72; DB 11; Length 211257;  
Best Local Similarity 68.9%; Pred. No. 1.1e-09;  
Matches 113; Conservative 0; Mismatches 50; Indels 1; Gaps 1;  
QY 278 ATCTGCAAGTCCCTGAGTGGCAGAGGAGCAGCCATGTGAGGCAAGAAATTCTGGC 337  
DB 72685 ATCTGCAAGCAGCCATGCGCATCAGATGAGATCTTGAGAGGCGCAAGATCTCTGGT 72626  
QY 338 TCAACACAGCTTAGCTCCCTGCTGTGTTCAACTTTGAGAGTTTGACCAAGCACTT 397  
DB 72625 TCAATTACAC-TAAATAGCTGAGAGTTGTTCAACATCGGAGAGCTGACTCCAAATAGTT 72567  
QY 398 TATTTTGGACATATTAAACAGACAGACAACTTTGGAAAAAGTT 441  
DB 72566 TATTTTGGACATATTAAAGCAGACATATTGTTGTAACAATGTT 72523  
RESULT 4  
ACN43912 ID ACN43912 standard; DNA; 233060 BP.  
XX  
XX ACN43912;  
XX 18-NOV-2004 (first entry)  
XX Mouse genomic sequence MCG11034.  
XX  
XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
XX  
XX Mus musculus.  
XX OS

XX  
PN WO2003073826-A2.  
XX 12-SEP-2003.  
XX 28-FEB-2003; 2003WO-US006235.  
XX 01-MAR-2002; 2002US-00087192.  
XX (SAGR-) SAGRES DISCOVERY.  
XX Morris DW;  
XX MPI; 2003-328604/31.  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX Claim 1; SEQ ID NO 97; Opp; English.  
XX The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (ix) as a biochip;  
XX carcinoma; (viii) for neutralizing the effect of CAP; and (xi) for  
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published  
XX  
SQ Sequence 233060 BP; 56260 A; 53970 C; 52927 G; 62345 T; 0 U; 7558 Other;  
Query Match 13.8%; Score 69.2; DB 11; Length 233060;  
Best Local Similarity 59.1%; Pred. No. 7.6e-09;  
Matches 137; Conservative 0; Mismatches 93; Indels 2; Gaps 1;  
QY 270 GTGGGATATCTGCAAGTCCATGATGTCAGAGGAGCAGCCATGTGAGGCAAGAAAGA 329  
DB 79572 GTAGAGTCCCTCCGAGCTCAGCACCACAGCAAGGAAAGCACTTGAGGTGAAGATGA 79631  
QY 330 ATCTGCTTCACACAGACTTAGCTCCCTGCTGTGTTCAACTTTGAGAGTTGACCAAC 389  
DB 79632 GCCCAGGTGACAGACACTTAGCAGAGGCCCTGG--GGCAGAAACCTCCAGAGTGTGACCTC 79689  
QY 390 AAGCAGCTTATTTTGTGACATATTAAACAGACAGCAACTTTGGAAAAAGTTTCTTATG 449  
DB 79690 AGGGTTTACCTTTCTGCTCAGCTTAAACAGTAAACCTTTGAGGAAAGTCTCCACGTG 79749  
QY 450 AAATTTATCAATTAAGCTTAAGCATGACTCATTAATGATCTTTGCA 501  
DB 79750 GCAATTATCAGAACCAAGCTTTTCAAGGATAGTTTCAATTAAGTCTCTTGCA 79801  
RESULT 5  
ABD32895 ID ABD32895 standard; DNA; 89824 BP.  
XX  
XX ABD32895;  
XX 18-NOV-2004 (first entry)  
XX Mouse cancer-associated genomic DNA MD18-037.  
XX  
XX Mouse; db; cancer-associated protein; gene; cytostatic; cancer;  
XX leukaemia; lymphoma; CAP.  
XX  
XX Mus musculus.  
XX OS

XX	MO2004074320-A2.
XX	02-SEP-2004.
XX	17-FEB-2004; 2004WO-US004730.
XX	14-FEB-2003; 2003US-00367094.
XX	14-MAR-2003; 2003US-00388838.
XX	15-APR-2003; 2003US-00417375.
XX	13-UN-2003; 2003US-00461862.
XX	15-SEP-2003; 2003US-00663431.
XX	15-DEC-2003; 2003US-00737318.
XX	(SAGR-) SAGRES DISCOVERY INC.
XX	
XX	Morris DW, Morris DW, Malandro MS;
XX	WPI; 2004-652914/63.
XX	
XX	New isolated cancer-associated polynucleotides and polypeptides useful
XX	for diagnosing, preventing or treating cancers, especially lymphoma and
XX	leukemia, or in screening for agents that modulate cancer.
XX	
XX	disclosure; seqid 591; 310pp; English.
XX	
XX	The invention relates to an isolated nucleic acid comprising at least 10
XX	contiguous nucleotides of any of the 233 polynucleotide sequences given
XX	in the specification, or its complement. The nucleic acids encode cancer-
XX	associated proteins. Also included are an expression vector comprising
XX	the isolated nucleic acid cited above, a host cell comprising the above
XX	recombinant nucleic acid or expression vector, a microarray for detecting
XX	a cancer-associated (CA) nucleic acid comprising at least one probe
XX	comprising at least 10 contiguous nucleotides of any of the above-
XX	mentioned nucleotide sequences, an isolated polypeptide (encoded within
XX	an open reading frame of a CA sequence selected from any of the 95
XX	polynucleotide sequences as mentioned in the specification, or its
XX	complement), an isolated antibody, (or its antigen binding fragment) that
XX	binds to the above polypeptide, a hybridoma that produces the above
XX	monoclonal antibody, a pharmaceutical composition comprising the above
XX	antibody and a pharmaceutical excipient, a kit for detecting cancer
XX	cells (comprising the antibody cited above, methods for diagnosing cancer
XX	or for detecting the presence or absence of cancer cells in an
XX	individual, a method for inhibiting growth of cancer cells in an
XX	individual, a method for delivering a therapeutic agent to cancer cells
XX	in an individual, an electronic library comprising the above
XX	polynucleotide or polypeptide (or their fragments), methods of screening
XX	for anticancer activity or for a bioactive agent capable of modulating
XX	the activity of a CA protein (CAP), methods for detecting cancer
XX	associated with expression of a polypeptide in a test cell sample, a
XX	method for treating cancers and a method for inhibiting the expression of
XX	CA gene in a cell. The composition and methods are useful for detecting,
XX	diagnosing, preventing and treating cancers, especially lymphoma and
XX	leukemia. These may also be used in screening for agents that modulate
XX	cancer. The present sequence is a mouse CAP genomic sequence. Note: The
XX	sequence data for this patent did not form part of the printed
XX	specification, but was obtained in electronic format directly from WIPO
XX	at ftp.wipo.int/pub/published_pot_sequences
XX	
XX	Sequence 89824 BP; 26577 A; 18285 C; 18254 G; 26575 T; 0 U; 133 Other;
XX	
XX	Query Match 13.1%; Score 65.6; DB 13; Length 89824;
XX	Best Local Similarity 68.5%; Pred. No. 5.9e-08;
XX	Matches 137; Conservative 0; Mismatches 54; Indels 9; Gaps 3
XX	
XX	272 GGGGATATCTGGCAAGTGGCCATGTAGTGGCAAGGACACGCCAATGTAGGCAAGAAGAAAT 3311
XX	DB 3139 GGGGATCTCTGCAAAATACATGTGTGCAAGAGTTTGACC-TTGAGAGGCTAATAAGACTTA 319
XX	332 TCTGGGCTCAACACAGCT-----TAGCTCCCTGGTGTGGTTCAAACTTTGAGAGTTTG 384
XX	DB 3198 TCGAGTTCAACAGAGACTCAGTAAGTAATATCAAGTCTTGTTCAAAATTTCTGAGAGCTG 325
XX	385 ACCCAAGACGACTTATTTTGTGACATATTTAAACAGACGACACACTTTGG-GAATAAGTTT 443

Db	3258	ACCTCAGTAGTAGTTAGTTAGGACCAATTAAAGCAGCACTAATTTGGAAAAAGTTTC	3317
Oy	444	CTTATGAAATATTCACAAAT	463
Db	3318	CTGATGCAATTGACTCAAT	3337

RESULT 6  
ABX16390 0/c

WP	Sequence split into 7 fragments	LOCUS	ABX16390	Accession	Abx16390
WP	Fragment Name	Begin	End		
WP	ABX16390_0	1	110000		
WP	ABX16390_1	100001	210000		
WP	ABX16390_2	200001	310000		
WP	ABX16390_3	300001	410000		
WP	ABX16390_4	400001	510000		
WP	ABX16390_5	500001	610000		
WP	ABX16390_6	600001	659158		
ID	ABX16390 standard; DNA; 659158 BP.				
AC	ABX16390;				
XX					
DT	09-APR-2003 (first entry)				
XX					
DE	Mouse high growth region.				
XX					
KM	High growth region; high growth phenotype; Soccs2; body size;				
KN	suppressor of cytokine signaling 2; ds; mouse.				
XX					
OS	Mus sp.				
XX					
FN	US2002155564-A1.				
XX					
PD	24-OCT-2002.				
XX					
PF	26-JAN-2001; 2001US-00771208.				
XX					
PR	29-DEC-1997; 97US-00999477.				
XX					
PA	(REGC ) UNIV CALIFORNIA.				
XX					
PI	Medrano JF, Bradford E, Horvat S;				
XX					
DR	WPI, 2003-182637/18.				
XX					
PT	Novel gene that when downregulated or knocked-out, results in high growth				
XX	phenotype; useful for regulating body size in mammals e.g. rodent, bovine				
XX	and canine.				
PS					
XX	Disclosure; SEQ ID NO 18; 49pp; English.				
XX					
CC	The invention describes an isolated nucleic acid molecule encoding a gene				
CC	product that, when knocked out, results in a high growth (hg) phenotype.				
CC	For example a nucleic acid disrupting the Soccs2 gene is useful for				
CC	producing an animal characterised by a hg phenotype, by inhibiting				
CC	expression of Soccs2 (suppressor of cytokine signaling 2) gene. The nucleic				
CC	acids of the invention are useful for regulating body size in mammals.				
CC	gene. The nucleic acids of the invention are useful for regulating body				
CC	size in mammals. This sequence represents the mouse high growth region.				
CC	Note: This sequence did not form part of the printed specification but				
CC	was obtained in electronic format directly from the US patent office at				
CC	seqdata.uspto.gov/sequence.html?DocID=20020155564				
XX					
XX					
Seq	Sequence 659158 BP; 177521A; 147222C; 149414G; 184754T; 0U; 2470bter;				
Query Match	13.0%; Score 65.2; DB 8; Length 110000;				
Best Local Similarity	73.8%; Pred. No. 8.5e-08;				
Matches	96; Conservative 0; Mismatches 33; Indels 1; Gaps 1;				
Oy	317	GAGGCAAGAAAGATTCCTGGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAACTTG	376		
Db	14437	GAGGCAAGAAAGATTCCTTAGTTCACACGACTCAG-TAGCTCATGTAGTTCAACTTG	14439		

QY 377 AGAGTTGACACACACACTTATTTTGCATATTTAAACAGACCAACTTTGGGAAA 436  
 DB 14438 AGAGTCTGACCTGAGAGTATTTTCAAGGATTTTAAAGCATATGACACTTCACTGAGAAA 14379  
 QY 437 AAGTTTCTT 446  
 DB 14378 AAAATTTCTT 14369

RESULT 7  
 ACN44860  
 ID ACN44860 standard, DNA; 48133 BP.  
 AC ACN44860;  
 XX 18-NOV-2004 (first entry)  
 DT  
 DE Mouse genomic sequence MCG49939.  
 XX  
 KM Cytosstatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO2003073826-A2.  
 PD 12-SEP-2003.  
 PF 28-FEB-2003; 2003MO-US006235.  
 XX  
 PR 01-MAR-2002; 2002US-00087192.  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 XX  
 DR WPI; 2003-328604/31.  
 XX  
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.  
 XX  
 PS Claim 1; SEQ ID NO 1519; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (ix) as a biochip;  
 CC carcinoma; (viii) for neutralizing the effect of CAP; and (xi) for  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xii) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US002182586A1, for which no sequence data was published  
 XX

SO Sequence 48133 BP; 13790 A; 10315 C; 10119 G; 13836 T; 0 U; 73 Other;

Query Match 12.7%; Score 63.6; DB 11; Length 48133;  
 Best Local Similarity 67.9%; Pred. No. 1.7e-07;  
 Matches 133; Conservative 0; Mismatches 59; Indels 4; Gaps 3;

QY 268 AGCGGGGATATTCGACAGTCCATGATGCGAGAGGAGCCCAATGTGAGGCAAG 327  
 DB 31444 ATGTAGAGGCTCTGTATTAACACAGACGACGAGCGTTAAGCC-ACATGAGTAAGAG 31502  
 QY 328 GAATCTGGCTCAACACAGCTTAGCTCCCTGGTGTGGTCAAACTTTGAGAGTTGACC 387  
 DB 31503 GAGCTGTGCTCAACACAGCTTAGTACC-GGTGTGTGTTTAACTTCAAGGCTCAACA 31561  
 QY 388 ACAAGCACTTATTTTGAATATTTAAACAGAGCAACAATTGGGAAAAAGTTTCTTA 447

DB 31562 CTGGGCAATTTATAGTATATTTTAAACACAGCAAA--TCTGAAAAATAGTTCTCG 31619  
 QY 448 TGAATAATATCACAAT 463  
 DB 31620 TGATAGTTAATCAAT 31635

RESULT 8  
 AB199246/c  
 ID AB199246 standard; cDNA; 2893 BP.  
 AC AB199246;  
 XX 07-MAR-2002 (first entry)  
 DT  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:72.  
 XX  
 KM Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200188188-A2.  
 PD 22-NOV-2001.  
 PF 18-MAY-2001; 2001WO-JP004192.  
 XX  
 PR 18-MAY-2000; 2000JP-00145977.  
 PA (UNIT-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 XX  
 PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 XX  
 DR WPI; 2002-034733/04.  
 XX  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.  
 XX  
 PS Claim 2; Page 225-227; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (II). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring the  
 CC expression levels of particular genes (AB199202 to AB199912, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. AB19913 and AB19914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention  
 XX

SO Sequence 2893 BP; 801 A; 607 C; 615 G; 870 T; 0 U; 0 Other;

Query Match 12.3%; Score 61.6; DB 6; Length 2893;  
 Best Local Similarity 65.0%; Pred. No. 1.9e-07;  
 Matches 91; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 362 TTGGTTCAACTTTGACAGTTTGACCAAGCACTTATTTTGAATATTTAAACAGAG 421  
 DB 2055 TGAATCAAGACTTTTGACACTGTATCTATGTTTATTTTCCACACATTTAAACATAG 1996  
 QY 422 CACAACCTTGGGAAAAAGTTTCTTATGAAATATATCAATTAAGCTTAAAGCACTGACT 481  
 DB 1995 TAGAATCTGGGAAAAAGTCTCTGTGACAACTTATTAACAAAGCTTACGGCCGCGCT 1936

QY 482 ACATTAATGCTTGGCA 501  
 DB 1935 ACATCTAAGTCCCTTGCA 1916

RESULT 9  
 ABD33090  
 ID ABD33090 standard; DNA; 145068 BP.

AC ABD33090;  
 DT 18-NOV-2004 (first entry)  
 DE Murine cancer-associated (CA) gene MD07-006.

XX Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;  
 KW ds; cancer; cytostatic.

OS Mus musculus.

PN WO2004058146-A2.

PD 15-JUL-2004.

PF 15-DEC-2003; 2003WO-US040081.

PR 17-DEC-2002; 2002US-00322281.

PS (SAGR-) SAGRES DISCOVERY INC.

PI Morris DW, Malandro MS;

DR WPI; 2004-499109/47.

XX Novel human cancer associated protein encoded within open reading frame  
 PT of cancer associated gene, useful as targets for diagnosing cancer.

PS Disclosure; SEQ ID NO 33; 182pp; English.

XX The invention relates to cancer-associated proteins (CAP) and the cancer-  
 CC associated (CA) nucleic acids encoding them. The invention also relates  
 CC to a method for treating cancers involving administering to a patient an  
 CC inhibitor of CAP, and a method of screening for anticancer activity in a  
 CC potential drug involving providing a cell that expresses a CA gene,  
 CC contacting a tissue sample derived from a cancer cell with an anticancer  
 CC drug candidate and monitoring the effect of the anticancer drug candidate  
 CC on expression of the CA gene. The CAP proteins are useful for detecting  
 CC cancer associated with expression of a CAP protein in a test cell sample  
 CC and for screening for a bioactive agent capable of modulating the  
 CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing  
 CC cancer, involving determining the expression of a CA nucleic acid in a  
 CC tissue. This sequence represents a murine CA gene of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 145068 BP; 38358 A; 33586 C; 33004 G; 36708 T; 0 U; 3412 Other;

QY Query Match 12.3%; Score 61.6; DB 13; Length 145068;  
 DB Best Local Similarity 58.3%; Pred. No. 1.1e-06;  
 Matches 126; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 275 GATATCTGCAATGCGATGAGGAGGACGCAATGAGGCAAGAAGAAATCTT 334

DB 101150 GGTCTCTGTCAGACACACAGCCAGCATGATATGATTCACAGAGGCAAGAACGAGTTG 101209

QY 335 GGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAACTTTGAGAGTTGACCAACAGCA 394

DB 101210 TGCTTCAATCAACTAAGTAGCAGCAGCTTGAGTCAAACTTTGGAGTGTGACCTCAGTA 101269

QY 395 CTTTATTTTGAATATTTAAACAGACAACTTTGGGAAAAAGTTTCTTATGAATAAT 454

DB 101270 GTTATTTTGAATATTTAAACATAGCACAATTTAATGMAAATTTCCAGTGTAACTT 101329

QY 455 TAT-CACATTAAGCTTAAGGATGACTACATTA 489  
 DB 101330 AATCCTGTATGTACAATAAGCATTAATCTCACCAAA 101365

RESULT 10  
 ADA02801/C  
 ID ADA02801 standard; DNA; 67832 BP.

AC ADA02801;

DT 06-NOV-2003 (first entry)

DE Mouse Itk carcinoma associated gene, SEQ ID NO:1319.

XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukemia; cytostatic; gene therapy; drug screening;  
 KW gene; ds.

OS Mus sp.

PN WO2003057146-A2.

PD 17-JUL-2003.

PF 26-DEC-2002; 2002WO-US041414.

PR 26-DEC-2001; 2001US-00035832.

PS (SAGR-) SAGRES DISCOVERY.

PI Morris DW;

DR WPI; 2003-587068/55.

XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.

PS Claim 1; SEQ ID NO 1319; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 67832 BP; 17054 A; 15504 C; 15824 G; 18278 T; 0 U; 1172 Other;

QY Query Match 12.3%; Score 61.4; DB 9; Length 67832;  
 DB Best Local Similarity 77.9%; Pred. No. 9.2e-07;  
 Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 407 CATATTTAAACAGAGCAAACTTTGGAAAAAGTTTCTTATGAATAATATCAATTA 466

DB 62076 CACATTTAAACAGATGAAGCTTTGGGAAAAAGTTTCTTGTGACAGTATCACGGCAAA 62017



QY 467 GCTTAAGCATGACTACATTAAATGCTTTGCAA 501  
 DB 62016 GCTTTAGCATGGCTACATCAAACTCTTTGTAA 61982

RESULT 11  
 ADB72539/c  
 ID ADB72539 standard; DNA; 67832 BP.

AC ADB72539;

DT 04-DEC-2003 (first entry)

DE Mouse Ick gene.

KM mouse; ds; cytosolic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

OS Mus sp.

PN MO2003008583-A2.

XX 30-JAN-2003.

PF 26-DEC-2001; 2001MO-US051291.

XX 02-MAR-2001; 2001US-00798586.

PR 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.

PR 30-NOV-2001; 2001US-00997722.

PR 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

PS Claim 1; SEQ ID NO 367; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytosolic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a mouse gene of the invention.

XX Sequence 67832 BP; 17054 A; 15504 C; 15824 G; 18278 T; 0 U; 1172 Other;

SO Query Match 12.3%; Score 61.4; DB 10; Length 67832;

Best Local Similarity 77.9%; Pred. No. 9.2e-07;

Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 407 CATATTTAAACAGACCAACTTTGGAAAAAGTTTCTTATGAAATTAATCAATATAA 466

DB 62076 CACATTTAAACAGACCAACTTTGGAAAAAGTTTCTTATGAAATTAATCAATATAA 62017

QY 467 GCTTAAGCATGACTACATTAAATGCTTTGCAA 501

DB 62016 GCTTTAGCATGGCTACATCAAACTCTTTGTAA 61982

RESULT 12  
 ADB85281/c

XX ID ADB85281 standard; DNA; 67832 BP.

XX AC ADB85281;

DT 01-JAN-2004 (first entry)

XX Mouse Ick genomic sequence.

XX Cytosolic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;  
 KW secreted; transmembrane; intracellular; ds.

OS Mus sp.

PN MO2003045230-A2.

XX 05-JUN-2003.

PF 02-DEC-2002; 2002MO-US038582.

PR 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

PT New recombinant nucleic acid comprising a nucleotide sequence of any of  
 PT the carcinoma-associated (CA) genes, useful for screening for drug  
 PT candidates for diagnosing or treating carcinomas.

PS Claim 1; SEQ ID NO 67; 983pp; English.

XX The invention relates to a recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the fully defined carcinoma-  
 CC associated (CA) genes from the 50 tables given in the specification. The  
 CC CA proteins are secreted, transmembrane or intracellular proteins. The  
 CC recombinant nucleic acids are useful for screening for drug candidates  
 CC for diagnosing or treating carcinomas. Sequences given in ADB85215-  
 CC ADB85514 represent CA genes of the invention.

XX Sequence 67832 BP; 17054 A; 15504 C; 15823 G; 18278 T; 0 U; 1173 Other;

SO Query Match 12.3%; Score 61.4; DB 10; Length 67832;

Best Local Similarity 77.9%; Pred. No. 9.2e-07;

Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 407 CATATTTAAACAGACCAACTTTGGAAAAAGTTTCTTATGAAATTAATCAATATAA 466

DB 62076 CACATTTAAACAGACCAACTTTGGAAAAAGTTTCTTATGAAATTAATCAATATAA 62017

QY 467 GCTTAAGCATGACTACATTAAATGCTTTGCAA 501

DB 62016 GCTTTAGCATGGCTACATCAAACTCTTTGTAA 61982

RESULT 13  
 ADM74396/c

XX ID ADM74396 standard; DNA; 67832 BP.

XX AC ADM74396;

XX 01-JUL-2004 (first entry)

DE Murine carcinoma associated (CA) nucleic acid #34.

XX Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;  
 KW carcinoma associated protein; CAP; carcinoma; leukemia; lymphoma;  
 KW cytosolic.

XX OS Mus musculus.

XX US2004072154-A1.

XX 15-APR-2004.

XX 30-NOV-2001; 2001US-00997722.



OS Mus musculus.  
XX WO2004058146-A2.  
XX  
XX 15-JUL-2004.  
PD  
XX 15-DEC-2003; 2003WO-US040081.  
PF  
XX 17-DEC-2002; 2002US-00322281.  
PR  
XX (SAGR-) SAGRES DISCOVERY INC.  
PA  
XX Morris DW, Malandro MS;  
PI  
XX WPI; 2004-499109/47.  
DR  
XX  
XX  
XX Novel human cancer associated protein encoded within open reading frame  
PT of cancer associated gene, useful as targets for diagnosing cancer.  
XX  
XX  
XX Disclosure; SEQ ID NO 587; 182bp; English.  
PS  
XX  
XX The invention relates to cancer-associated proteins (CAP) and the cancer-  
CC associated (CA) nucleic acids encoding them. The invention also relates  
CC to a method for treating cancers involving administering to a patient an  
CC inhibitor of CAP, and a method of screening for anticancer activity in a  
CC potential drug involving providing a cell that expresses a CA gene,  
CC contacting a tissue sample derived from a cancer cell with an anticancer  
CC drug candidate and monitoring the effect of the anticancer drug candidate  
CC on expression of the CA gene. The CAP proteins are useful for detecting  
CC cancer associated with expression of a CAP protein in a test cell sample  
CC and for screening for a bioactive agent capable of modulating the  
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing  
CC cancer, involving determining the expression of a CA nucleic acid in a  
CC tissue. This sequence represents a murine CA gene of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX  
SQ Sequence 121129 BP; 34090 A; 24248 C; 25630 G; 36418 T; 0 U; 743 Other;  
  
Query Match 12.1%; Score 60.4; DB 13; Length 121129;  
Best Local Similarity 71.3%; Pred. No. 2.4e-06;  
Matches 107; Conservative 0; Mismatches 41; Indels 2; Gaps 2;  
  
QY 314 TGTAGGCGAAGAGATTCGCTCAACAGCTTAGCTCCTGTGTGTGTTCAACT 373  
DB 119057 TGAGAGCAAAAGGCGAGTTGTTCAATGACTCA-ATAGCTGTGTAGTTCAATT 118999  
QY 374 TTGAGAGTTGACCAAGCACTTTATTTTGTGACATATTTAAACAGCAACTTTGGG 433  
DB 118998 TTGAAAGCTGACCTCAAGTATTTTATTTAGGCAATTTAGCAGCAGAA-ATTGAA 118940  
QY 434 AAAAGTTTCTATGAAATTTATCAAT 463  
DB 118939 AAGAAAGTTGTAAATGATTAATAGCTTAAT 118910

Search completed: February 10, 2005, 09:41:03  
Job time : 345.632 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58, Search time 103.92 Seconds  
(without alignment)  
7888.487 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_5000\_5500

Perfect score: 501  
Sequence: 1 ggtcagccatgtcttattt.....acataaataatgcttgcga 501

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCBUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	7.5	144362	4 US-09-949-016-16066	Sequence 16066, A
2	35.8	7.1	15027	4 US-09-949-016-12660	Sequence 12660, A
3	35.8	7.1	15036	4 US-09-949-016-13351	Sequence 13351, A
4	35.6	7.1	1053	4 US-09-710-279-3127	Sequence 3127, Ap
5	35.6	7.1	1149	3 US-09-134-001C-2612	Sequence 2612, Ap
6	35.6	7.1	1492	4 US-09-710-279-3677	Sequence 3677, Ap
7	35.6	7.1	3825	4 US-09-710-279-3794	Sequence 3794, Ap
8	34.8	6.9	7218	1 US-08-232-463-14	Sequence 14, Appl
9	34.8	6.9	10543	4 US-09-949-016-13138	Sequence 13138, A
10	34.4	6.9	101472	4 US-09-949-016-15861	Sequence 15861, A
11	34.4	6.9	102738	4 US-09-949-016-12447	Sequence 12447, A
12	34.4	6.9	192700	4 US-09-949-016-11820	Sequence 11820, A
13	34.4	6.9	192704	4 US-09-949-016-17182	Sequence 17182, A
14	34.2	6.8	600	4 US-09-621-976-14175	Sequence 14175, A
15	34	6.8	118382	4 US-09-949-016-15996	Sequence 15996, A
16	34	6.8	118382	4 US-09-949-016-15997	Sequence 15997, A
17	33.8	6.7	1141	4 US-09-806-708B-22	Sequence 22, Appl
18	33.8	6.7	86439	4 US-09-949-016-11945	Sequence 11945, A
19	33.8	6.7	86440	4 US-09-949-016-16990	Sequence 16990, A
20	33.6	6.7	7005	4 US-09-949-016-2891	Sequence 2891, Ap
21	33.6	6.7	7005	4 US-09-949-016-2892	Sequence 2892, Ap
22	33.6	6.7	7005	4 US-09-949-016-2893	Sequence 2893, Ap
23	33.6	6.7	7005	4 US-09-949-016-2894	Sequence 2894, Ap
24	33.6	6.7	57280	4 US-09-949-016-11796	Sequence 11796, A
25	33.6	6.7	57280	4 US-09-949-016-12843	Sequence 12843, A
26	33.6	6.7	57280	4 US-09-949-016-12844	Sequence 12844, A
27	33.6	6.7	57280	4 US-09-949-016-12846	Sequence 12846, A

28	33.6	6.7	57280	4 US-09-949-016-13542	Sequence 13542, A
29	33.6	6.7	57280	4 US-09-949-016-13543	Sequence 13543, A
30	33.6	6.7	57280	4 US-09-949-016-13544	Sequence 13544, A
31	33.6	6.7	57280	4 US-09-949-016-13545	Sequence 13545, A
32	33.6	6.7	57280	4 US-09-949-016-14633	Sequence 14633, A
33	33.6	6.7	57280	4 US-09-949-016-14634	Sequence 14634, A
34	33.6	6.7	57280	4 US-09-949-016-14635	Sequence 14635, A
35	33.6	6.7	57280	4 US-09-949-016-14636	Sequence 14636, A
36	33.6	6.7	57280	4 US-09-949-016-14637	Sequence 14637, A
37	33.6	6.7	57280	4 US-09-949-016-14638	Sequence 14638, A
38	33.6	6.7	57280	4 US-09-949-016-14639	Sequence 14639, A
39	33.6	6.7	57280	4 US-09-949-016-14640	Sequence 14640, A
40	33.6	6.7	96866	4 US-09-949-016-13768	Sequence 13768, A
41	33.6	6.7	116592	4 US-09-818-512-3	Sequence 3, Appl1
42	33.6	6.7	152132	4 US-09-949-016-13845	Sequence 13845, A
43	33.6	6.7	152145	4 US-09-949-016-12371	Sequence 12371, A
44	33.4	6.7	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
45	33.4	6.7	1664976	4 US-09-692-570-1	Sequence 1, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-949-016-16066/c
; Sequence 16066, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 16066
; LENGTH: 144362
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(144362)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16066

Query Match 7.5%; Score 37.6; DB 4; Length 144362;
Best Local Similarity 54.5%; Pred. No. 2.1;
Matches 96; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 81 AGTCCAGGCTGACAGTGTGATACACAGGACGATGAGGCTCCACGCTGAACAG 140
DB 12004 AGTCTATCTCTGAGATGATGACAGAAAGAGGAGGACAGTCTCCAGAGGAA 11945
QY 141 TCAAGCTGGAGAGAGAAAGACGACGACATCTCTCAACCAATGATGCTTGA 200
DB 11944 ACGAGCTCAGAAAG-GAAGAAAGGAGCAAAAGAGATGATGATTTTCCCAAT 11886
QY 201 CAACATATATATATATATATATATATATATATATATATATATATATATATAT 256
DB 11885 AACCTACATATATATATATATATATATATATATATATATATATATATATAT 11830

RESULT 2
US-09-949-016-12660
; Sequence 12660, Application US/09949016
; Patent No. 6812339
```

Query Match	7.1%	Score 35.8	DB 4	Length 15036
Best Local Similarity	50.3%	Pred. No. 2.1		
Matches	88	Conservative	0	Mismatches 87; Indels 0; Gaps 0;
Qy	156	AGAAAGACAGACACATTCCTTCACCAACTATGTCTGAAAAACAAACATATTATATC	215	
Db	1331	AGAAATTCAGGCCCTCATATGCCCAAGCACTGTATCTAGAAAAATATCTCAATGGTAAG	1339	
Qy	216	ACATATATTCGATTTATAGACAGCTAAATATCTGGGTAGCATATCTCCAGGTGGG	275	

Query Match	7.1%	Score 35.6	DB 3	length 1149
Best Local Similarity	55.7%	Pred No. 0.55		
Matches 68; Conservative	0	Mismatches 54	Indels 0	Gaps 0

OY	369	AAACTTTGAGAGTTTGACCAACAAGCACTTATTTTGGACATATTTTAAACAGACCAACT	428
Db	497	AATCTTTCACAGTTGAATCTGAAGAATATATGATTTAAGTCAAATTCGAACGAGTCAATTA	556
OY	429	TTGGGAAAAGTTTCTTATGAAAATATCAATATAAGCTTAAGGACATGACATTA	488
Db	557	TTACCGTAGATGTTTCTTAGTGAAGAAAGTTACCATATAGAGCTAAACGTCATTATCAATCA	616
OY	489	AA 490	
Db	617	TA 618	

  

RESULT 6			
US-09-710-279-3677/c			
Sequence 3677, Application US/09710279			
Patent No. 6703492			
GENERAL INFORMATION:			
APPLICANT: KIMBERLY, WILLIAM JOHN			
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS			
FILE REFERENCE: PU3480US			
CURRENT APPLICATION NUMBER: US/09/710,279			
CURRENT FILING DATE: 2000-11-09			
PRIOR APPLICATION NUMBER: 60/164,258			
PRIOR FILING DATE: 1999-11-09			
NUMBER OF SEQ ID NOS: 4472			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 3677			
LENGTH: 1492			
TYPE: DNA			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: synthetic			
OTHER INFORMATION: nucleic acid sequence			
US-09-710-279-3677			

  

Query Match	7.1%;	Score 35.6;	DB 4;	Length 1492;
Best Local Similarity	55.7%;	Pred. No. 0.64;		
Matches	68;	Conservative	0;	Mismatches 54; Indels 0; Gaps 0;

  

OY	369	AAACTTTGAGAGTTTGACCAACAAGCACTTATTTTGGACATATTTTAAACAGACCAACT	428
Db	1411	AATCTTTCACAGTTGAATCTGAAGAATATGATTTAAGTCAAATTCGAACGAGTCAATTA	1352
OY	429	TTGGGAAAAGTTTCTTATGAAAATATCAATATAAGCTTAAGGACATGACATTA	488
Db	1351	TTACCGTAGATGTTTCTTAGTGAAGAAAGTTACCATATAGAGCTAAACGTCATTATCAATCA	1292
OY	489	AA 490	
Db	1291	TA 1290	

  

RESULT 7			
US-09-710-279-3794			
Sequence 3794, Application US/09710279			
Patent No. 6703492			
GENERAL INFORMATION:			
APPLICANT: KIMBERLY, WILLIAM JOHN			
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS			
FILE REFERENCE: PU3480US			
CURRENT APPLICATION NUMBER: US/09/710,279			
CURRENT FILING DATE: 2000-11-09			
PRIOR APPLICATION NUMBER: 60/164,258			
PRIOR FILING DATE: 1999-11-09			
NUMBER OF SEQ ID NOS: 4472			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 3794			
LENGTH: 3825			
TYPE: DNA			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: synthetic			

```

; OTHER INFORMATION: nucleic acid sequence
; US-09-710-279-3794

Query Match          7.1%; Score 35.6; DB 4; Length 3825;
Best Local Similarity 5.7%; Pred. No. 1.1;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Cy 369 AAACCTTGAGAGTTTGACCAACAAGCACTTATTTTGGACATATTAAACGAGCAACT 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1922 AATCTTCACAGTTGAATCTGAAGAATATGATTAAAGTCAATTCGAAACCGGTCATACA 1981

Cy 429 TTGGGAAAAAGTTTCTTATGAAAAATATCAATAAAGTTTAAGCATGACATCAATTAA 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1982 TTACCGTAGATGTTTCTTAGTGAAGAAAAGTTACCATATGAGCTAAACGTCATTATTCATCA 2041

Cy 489 AA 490
    |
Db 2042 TA 2043

RESULT 8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ5pT-Fls
; US-08-232-463-14

Query Match          6.9%; Score 34.8; DB 1; Length 7218;
Best Local Similarity 5.7%; Pred. No. 2.9;
Matches 18; Conservative 164; Mismatches 136; Indels 0; Gaps 0;

```

```

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15861
; LENGTH: 101472
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15861

Query Match          6.9%; Score 34.4; DB 4; Length 101472;
Best Local Similarity 56.0%; Pred. No. 18;
Matches 65; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY      357 TGGTGTGGTTCAAACTTTGAGAGTTGACCAAGCAACTTATTTTGCATATTTAA 416
DB      95699 TTGCTTTTAAACTCTCTAAATCTTAACTCATGCCATTTTATTTATAAAGTATA 95640

QY      417 CAGAGCAACAATTGGGAAAAAGTTTCTTAGAAAATATATCAATAAGCTTAA 472
DB      95639 AATAAACTACTTCCAAATCATCTGTTTAAACAATTATTAATTAATTTA 95584

RESULT 11
US-09-949-016-12447/c
; Sequence 12447, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12447
; LENGTH: 102738
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12447

Query Match          6.9%; Score 34.4; DB 4; Length 102738;
Best Local Similarity 56.0%; Pred. No. 18;
Matches 65; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY      357 TGGTGTGGTTCAAACTTTGAGAGTTGACCAAGCAACTTATTTTGCATATTTAA 416
DB      95750 TTGCTTTTAAACTCTCTAAATCTTAAACTCATGCCATTTTATTTATAAAGTATA 95691

QY      417 CAGAGCAACAATTGGGAAAAAGTTTCTTAGAAAATATATCAATAAGCTTAA 472
DB      95699 AATAAACTACTTCCAAATCATCTGTTTAAACAATTATTAATTAATTTA 95635

RESULT 12
US-09-949-016-11820/c
; Sequence 11820, Application US/09949016

```



```

1 Patent No. 6812339
2 ;
3 GENERAL INFORMATION:
4 ;
5 APPLICANT: VENTER, J. Craig et al.
6 ;
7 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
8 ;
9 TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
10 ;
11 FILE REFERENCE: CLO01307
12 ;
13 CURRENT FILING DATE: US/09/949,016
14 ;
15 CURRENT FILING DATE: 2000-04-14
16 ;
17 PRIOR APPLICATION NUMBER: 60/241,755
18 ;
19 PRIOR FILING DATE: 2000-10-20
20 ;
21 PRIOR APPLICATION NUMBER: 60/237,768
22 ;
23 PRIOR FILING DATE: 2000-10-03
24 ;
25 PRIOR APPLICATION NUMBER: 60/231,498
26 ;
27 PRIOR FILING DATE: 2000-09-08
28 ;
29 NUMBER OF SEQ ID NOS: 207012
30 ;
31 SOFTWARE: FASTSEQ for Windows Version 4.0
32 ;
33 SEQ ID NO 11820
34 ;
35 LENGTH: 192700
36 ;
37 TYPE: DNA
38 ;
39 ORGANISM: Human
40 ;
41 US-09-949-016-11820

```

	Query Match	6.9%	Score 34.4;	DB 4;	Length 192700;	
	Best Local Similarity	63.1%;	Pred. No. 27;	Mismatches	31; Indels	0; Gaps 0;
	Matches	53; Conservative	0;			
Oy	163 CCAGACACATCTTTCACAACAATGCTGTGGAAAAAACAACATATTATATCATAATA	222				
Dd	167541 CCACATCACACTCCCTTAAGTGCGGCCCATGCGCTTAGAACCAATGACTATATTAAATACCTATA	167482				
Oy	223 TTGCATTTATGAGACAGCTAAAT	246				
Dd	167481 TTTAATTATCATTTTGGTTAAAT	167458				

```

RESULT 13
US-09-949-016-17182/c
; Sequence 17182, Application US//09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17182
; LENGTH: 192704
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17182

```

	Query Match	6.9%	Score 34.4;	DB 4;	Length 192704;
	Best Local Similarity	63.1%;	Pred.	No. 27;	
	Matches	53; Conservative	0;	Mismatches	31; Indels
				Gaps	0;
Oy	163 CCAGCACATCTCCTCAACCAACTAAGTGTGGAATAAACAACATATTATTCACATATA	222			
Dd	167541 CCGATGCACCTCCCTTAACCTGGGCCCATGCTTGACCATGACATTTATTATCACTATA	16748			
Oy	223 TTGCATTATGAGACAGACTAAAAT	246			
Dd	167481 TTTAATTATCATTTTGGTTAAAT	167458			

```

RESULT 14
US-09-621-976-14175/c
; Sequence 14175, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J. B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J. Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14175
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-14175

```

Query Match	Best Local Similarity	6.8%;	Score 34.2;	DB 4;	Length 600;				
Matches	87;	Conservative	0;	Mismatches	88;	Indels	0;	Gaps	0;
QY	280	CTGCAAGTCCANTAGTGGCAGAGGGAACGCCAATGTGAGGCAAGAGGAATTTCTGCTC	339						
DB	494	CTGCACAGGATTTTCAGAGGAGAGGAGATGATGTGAGGCAACAGGATTTTTCAGGGC	435						
QY	340	AACACAGTTAGTCCCTCGTGTGTTGTTCAAATTGAGAGTTTGACACAGCACTTTA	399						
DB	434	AGTGAAGCTGACGTTACAGTGTGGGTACATCTCTATCAATTTGTCAATGTATCTTTGA	375						
QY	400	TTTTGCATATTTAAACAGAGCACACTTTGGGAAAAAGTTTCTTATGAAAAAT	454						
DB	374	ATATACACAAAGACAAAACCTAACAACTCCGACCTTTGGCTATATATATATAT	320						

```

RESULT 15
US-09-949-016-15996
; Sequence 15996, Application US/09949016
; Patent No. 6812319
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15996
; LENGTH: 118382
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(118382)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15996

```

Query Match	6.8%	Score 34	DB 4	Length 118382
Best Local Similarity	51.3%	Pred. No. 27		
Matches	79	Conservative	0	Mismatches 75; Indels 0; Gaps 0

  

Qy	112	GCAGATGAGGCTCTCTCAGCTGAAAGCAGTCAAGCTGGCAGAAAGAAAGACACAC	171
Db	79717	GCACCTTAGAGGCCCAAGCAGAGAGATTTGTTGAGACCAAGAGTTTGGACACACCTGG	7977

OY 172 ATTCTTCAACCACTAGTCTTGAAACAACATATTATATCACATATATGCAATTTA 231  
Db 79777 GTAACATAGACCTTGTCTTTATTAATAAATAAGTAATTAAGTACATGTATATTAAGA 79836  
OY 232 TGAGACAGCTAAATGTACTCGGCTAGCATGACT 265  
Db 79837 TTAAGGAATATATATTACAGTGTAGCAGTAAT 79870

Search completed: February 10, 2005, 17:42:49  
Job time : 107.92 secs



Db 72685 ATTCGACCCACCATGCGCATCAGTAAGATCTTGAAGAGGACAGAGATACCTCTGCT 72626  
QY 338 TCACACAGCTTAGCTCCCTGGTGTGTTGTTCAAACTTTGAGATTGACCAAGACTT 397  
Db 72625 TCACATACAC- TAATAGCTGAGAGTTGGTCAAAATCATCGGAGGTCTGACTCCAAATAGTT 72567  
QY 398 TATTTTGAATATTTTAAACAGACAGCAACTTTGGGAAAAAGTT 441  
Db 72566 TATTTTGAATATTTTAAAGCAGCAATATTTGGTGAATAGTT 72523

RESULT 2  
US-10-087-192-97  
; Sequence 97, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; PRIOR FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 97  
; LENGTH: 233060  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(233060)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-97

Query Match 13.8%; Score 69.2; DB 13; Length 233060;  
Best Local Similarity 59.1%; Pred. No. 1.9e-08;  
Matches 137; Conservative 0; Mismatches 93; Indels 2; Gaps 1;

QY 270 GTGGGATATCTGCAAGTGCATGATGCGACAGAGGACGCCAATGTGAGGCAAGAGA 329  
Db 79572 GTGAGGGTCCCTGCGCCTCAACACAGGAAAGCCACTGAGGTGAAGATGA 79631  
QY 330 ATTCTGCTCAACAGACTTAGCTCCCTGGTGTGTTCAAACTTTGAGATTGACAC 389  
Db 79632 GCCCAGGTCAAGCACTTAGAGAGCCCTGG--GGCAGAAACCTCCAGATCTGACCTC 79689  
QY 390 AAGCATTATTTTGAATATTTAAACAGACAACTTTGGGAAAAAGTTTCTTATG 449  
Db 79690 AAGGTTACCTTCTGCTCACTTAAACAGTAAACCTTTGAGGAAAGTCTCCACG 79749  
QY 450 AAAATATATCAATTAAGCTTAAGCACTGACTAATTAATTCCTTTGAA 501  
Db 79750 GCAATATACGAACAGCTTACAGGATAGTTTCATTGAACCTCCCTTGAA 79801

RESULT 3  
US-09-771-208-20/c  
; Sequence 20, Application US/09771208  
; Patent No. US2002015564A1  
; GENERAL INFORMATION:  
; APPLICANT: MEDRANO, JUAN  
; APPLICANT: BRADFORD, ERIC  
; APPLICANT: HORVAT, SIMON  
; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE  
; FILE REFERENCE: 407T-923710US  
; CURRENT APPLICATION NUMBER: US/09/771,208  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: US 08/999,477

PRIOR FILING DATE: 1997-12-29  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 20  
LENGTH: 659158  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (123459)..(123478)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (602466)..(602485)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (546986)..(547017)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (494715)..(494814)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (390986)..(391005)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (346860)..(346823)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (317174)..(317193)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (280353)..(280373)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (271829)..(271848)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (183872)..(183891)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (170625)..(170645)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
LOCATION: (132680)..(132700)  
OTHER INFORMATION: n is unidentified a, c, g, or t  
NAME/KEY: misc.feature  
OTHER INFORMATION: n is a, c, g, or t  
US-09-771-208-20

Query Match 13.0%; Score 65.2; DB 9; Length 659158;  
Best Local Similarity 73.8%; Pred. No. 5.2e-07;  
Matches 96; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 317 GAGGCAAGAGAAATTTGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAAACTTTG 376  
Db 14497 GAGGCAAGAAAGAAAGCTGATTCACACAGCTCAG-TAGCTATGTATTTAACTTTG 14439  
QY 377 AAGATTGACCAAGCACTTATTTTGAATATTTTAAACAGACAACTTTGGGAAA 436  
Db 14438 AAGATCTGACCTTGAGTATTTTCAAGGATTTTAAGCATAGACACTTCAGTGAAA 14379  
QY 437 AAGTTTCTT 446  
Db 14378 AAAATTTCTT 14369

RESULT 4  
US-10-175-523-57/c  
; Sequence 57, Application US/10175523  
; Publication No. US20030096264A1  
; GENERAL INFORMATION:  
; APPLICANT: Brockman, Jeffrey  
; APPLICANT: Evans, David  
; APPLICANT: Hook, Derek

```
APPLICANT: Klimczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Paleyeman, Michael
APPLICANT: Rajan, Pritih
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
FILE REFERENCE: 3235/11795-US3
CURRENT APPLICATION NUMBER: US/10/175,523
CURRENT FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/299,151
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/317,828
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 57
LENGTH: 225883
TYPE: DNA
ORGANISM: Mus musculus
US-10-175-523-57

Query Match      12.8%; Score 64.2; DB 14; Length 225883;
Best Local Similarity 61.7%; Pred. No. 5.8e-07;
Matches 119; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY      271 TGGGGATATCTGCAGAGTCCATGATGTCGAGGAGGACAGCCCATGTGAGGCAAGAGAA 330
DB      215423 TGGAGGCTCTGTGACCTTCAAGCCCATCAAAAAGATGATCCCAAGAGCTGAGAAAGAA 215364
QY      331 TTCTGCTCAACACAGCTTAGCTCCCTGGCTTGGTTCAACTTGGAGTTGACACA 390
DB      215363 ATTGTTCACACACACTGCA-ATGGCTGGTGTGATTTAAATTTCTGAGCTGACCTCCA 215305
QY      391 AGCATTATTTTGGACATATTTTAAACAGACACACTTTGGGAAAAAGTTTCTTATGA 450
DB      215304 AGGAGTTATTTTGGGAGGTTTAAAGCAGCCCAATTTGCTCAAACTTACTACTAGTGA 215245
QY      451 AAATTATCACAAT 463
DB      215244 AAATTACTCACT 215232

RESULT 5
US-10-087-192-1519
; Sequence 1519, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1519
; LENGTH: 48133
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
```

```
LOCATION: (1)...(48133)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1519

Query Match      12.7%; Score 63.6; DB 13; Length 48133;
Best Local Similarity 67.9%; Pred. No. 3.9e-07;
Matches 133; Conservative 0; Mismatches 59; Indels 4; Gaps 3;

QY      268 AGTGGGATATCTGCAAGTCATGAGTGGGACAGGAGGACAGCCCATGTGAGGCAAGAG 327
DB      31444 ATGTGAGGAGCTCTGTACACACAGCAGCAGGCGTTAAGCC-ACATGAGTAAAGAG 31502
QY      328 GAATTCGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAAACTTGGAGAGTTGACACC 387
DB      31503 GAAGTCTGCTCAACACACTTAGTAGCC-GGTGTGTTTAAACTTCAAGGGTCCACA 31561
QY      388 ACAAGACCTTATTTTGAATTTTAAACAGACCAACTTTGGGAAAAAGTTTCTTA 447
DB      31562 CTGGGCAATTATGTAGTATATTTAGCACAACAACA--TCTGAAAAATAGTTTCTTG 31619
QY      448 TGAATATATCACAAT 463
DB      31620 TGATAGTTAATCAAT 31635

RESULT 6
US-10-085-117-121/C
; Sequence 121, Application US/10085117
; Publication No. US2003023334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121
; LENGTH: 47884
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(47884)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-121

Query Match      12.5%; Score 62.4; DB 17; Length 47884;
Best Local Similarity 57.5%; Pred. No. 8.8e-07;
Matches 131; Conservative 0; Mismatches 96; Indels 1; Gaps 1;

QY      270 GTGGGATATCTGCAAGTCATGAGTGGGACAGGAGGACAGCCCATGTGAGGCAAGAG 329
DB      40174 GTGGGATCTCTGAAACACCTTAAACAGAGGCTGTGATTCAGAGAGAGAGAGAA 40115
QY      330 ATTCTGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAAACTTGGAGAGTTGACCA 389
DB      40114 AGTGTCTCAAGTGAAGTCAAGTGAAGTTT-GTCCAAACTTGGAGTGAAGCACC 40056
QY      390 AAGCACTTATTTTGAATTTTAAACAGACCAACTTTGGGAAAAAGTTTCTTATG 449
DB      40055 AAGTAGTTATTTTGAAGCAGCAGCAATTTCAAAATGTTCTGATATATTAATCACT 39996
QY      450 AAATTATCAATTAAGCTTAAGGATGATGATCAATTAATGCTTT 497
DB      39995 CCTGTGTACAGAAAGCTTAGGAGATGGCTATCAAACTTTGT 39948

RESULT 7
US-10-417-375-15
```

```
/ Sequence 15, Application US/10417375
/ Publication No. US20040219528A1
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/417,375
/ CURRENT FILING DATE: 2003-04-15
/ NUMBER OF SEQ ID NOS: 176
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15
/ LENGTH: 105077
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)-(105077)
/ OTHER INFORMATION: n = A,T,C or G
US-10-417-375-15
```

```
Query Match          12.3%; Score 61.8; DB 18; Length 105077;
Best Local Similarity 62.2%; Pred. No. 2e-06;
Matches 115; Conservative 0; Mismatches 67; Indels 3; Gaps 1;
```

```
QY 317 GAGGCAAGAGAAATTCGCTCACAAGCTTACCTCCCTGGTGGTTCAACTTGG 376
DB 46940 GTGTGAAGATGAACTGGGCTGACGACAC--GTTACAGACGTTCTCAAAACCTTC 46996
QY 377 AAGATTGACCAAGACCTTTATTTGACATTTAAAGACAGCAACTTTGGGAAA 436
DB 46997 AGAGTCTACCTCAGCTATTCTTTGACACATTTAAGTGAAGTAAACGTTGCAAG 47056
QY 437 AAGTTTCTTATGAAATTAATCAATTAAGCTTAAAGCATGACTTAATAATGCTT 496
DB 47057 AGTTTCCACATGACATTAGCAACACTATGTTTAGATAGCAGATAGAAACTCCCC 47116
QY 497 TGCAA 501
DB 47117 AGCAA 47121
```

```
RESULT 8
US-10-322-281-33
/ Sequence 33, Application US/10322281
/ Publication No. US20040126762A1
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc S. Malandro
/ TITLE OF INVENTION: Novel Compositions and Methods in Cancer
/ FILE REFERENCE: 529452001000
/ CURRENT APPLICATION NUMBER: US/10/322,281
/ CURRENT FILING DATE: 2002-12-17
/ NUMBER OF SEQ ID NOS: 866
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 33
/ LENGTH: 145068
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)-(145068)
/ OTHER INFORMATION: n = A,T,C or G
US-10-322-281-33
```

```
Query Match          12.3%; Score 61.6; DB 18; Length 145068;
Best Local Similarity 58.3%; Pred. No. 2.7e-06;
Matches 126; Conservative 0; Mismatches 89; Indels 1; Gaps 1;
```

```
QY 275 GATATCTGCAAGTGCATGAGTGACAGAGGACCAATGTAGGCAAGAAATTTCT 334
DB 101150 GGTCCTGTGACGACGACGACGACGACGATGTAATGATCCACAGAGGCAAGACGAGTTG 101209
```

```
QY 335 GGCTCAACAGACTTAGCTCCCTGGTGGTTCAAACTTTGAGAGTTTGACCAAGCA 394
DB 101210 TGGTCAATACAACTAGTACGACGCTGGCTAAACTTTGGAGTGTGACCTCAGTA 101269
QY 395 CTTTATTTTGAATTTTAAACAGACACAACTTTGGGAAAAAGTTTCTTATGAAAT 454
DB 101270 GTTATTTTAGGTATGTTTAAACATAGCACAATTTAGTGAATAATTTCCAGTGGTAATT 101329
QY 455 TAT-CACAAATAAGCTTAAGGCAATGACTCAATTAAA 489
DB 101330 AATCCTGTATGTACAAATAGGCAATTAATCCACCAA 101365
```

```
RESULT 9
US-09-997-722-67/C
/ Sequence 67, Application US/09997722
/ Publication No. US20040072154A1
/ GENERAL INFORMATION:
/ APPLICANT: Morris, David
/ APPLICANT: Engelhard, Eric
/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
/ FILE REFERENCE: A-71171/RMS/DCF
/ CURRENT APPLICATION NUMBER: US/09/997,722
/ CURRENT FILING DATE: 2001-11-30
/ PRIOR APPLICATION NUMBER: US 09/747,377
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 09/798,586
/ PRIOR FILING DATE: 2001-03-02
/ SOFTWARE: PatentIn version 3.1
/ NUMBER OF SEQ ID NOS: 301
/ SEQ ID NO 67
/ LENGTH: 67832
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (28892)-(29324)
/ OTHER INFORMATION: "n" at positions 28892 through 29324 can be any base.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (30179)-(30757)
/ OTHER INFORMATION: "n" at positions 30179 through 30757 can be any base.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (32761)-(32780)
/ OTHER INFORMATION: "n" at positions 32761 through 32780 can be any base.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (43840)-(43859)
/ OTHER INFORMATION: "n" at positions 43840 through 43859 can be any base.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (57867)-(57886)
/ OTHER INFORMATION: "n" at positions 57867 through 57886 can be any base.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (60656)-(60675)
/ OTHER INFORMATION: "n" at positions 60656 through 60675 can be any base.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (61916)-(61935)
/ OTHER INFORMATION: "n" at positions 61916 through 61935 can be any base.
US-09-997-722-67
```

```
Query Match          12.3%; Score 61.4; DB 11; Length 67832;
Best Local Similarity 77.9%; Pred. No. 2.1e-06;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 407 CATATTTAAACAGACAACTTTGGGAAAAAGTTTCTTATGAAATTAATCAATATA 466
```

Db 62076 CACATTAAACACAGTAAGCTTTGGGAAAAGTTCTTGTCACAGTATGATCGGCAAA 62017  
QY 467 GCTTAAGCATGACTACATTAATAATGCTTTGCAA 501  
62016 GCTTTAGGCATGGCTACATCAAAACTTTTGTAA 61982

RESULT 10  
US-10-087-192-289  
; Sequence 289, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 5294520012  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 289  
; LENGTH: 210528  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(210528)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-289

Query Match 12.1%; Score 60.8; DB 13; Length 210528;  
Best Local Similarity 63.4%; Pred. No. 5.8e-06;  
Matches 109; Conservative 0; Mismatches 62; Indels 1; Gaps 1;  
QY 284 AAGCGCATGATGCGGAGAGGAGACAGCAATGTGAGGAGAGAAATTCGCTCAACA 343  
Db 172729 AATCTCTGCTGAGTACAGAGTAACACTTACAGAGAGAGAGAGCCAGTTGCTTCAAC 172788  
QY 344 CAGCTTAGCTCCCTGCTGTTGTTCAACTTTGAGAGTTGACCAACAAGCACTTATTTT 403  
Db 172789 CAATCTCA-CTAGTGAAGTGTGTTCAAGCTTCTGGGGTCTGACCCCTAGTAGTTACTAT 172847  
QY 404 TGACATATTTAAACAGACACACACTTTGGGAAAAAGTTTCTTATGAAAT 455  
Db 172848 AGGACATTTAAAGTACAGTACATTTGATGAAAGATGATCTGTGCTAAT 172899

RESULT 11  
US-10-322-281-587/c  
; Sequence 587, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 52945200100  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 587  
; LENGTH: 121129  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(121129)

OTHER INFORMATION: n = A,T,C or G  
US-10-322-281-587  
Query Match 12.1%; Score 60.4; DB 18; Length 121129;  
Best Local Similarity 71.3%; Pred. No. 5.7e-06;  
Matches 107; Conservative 0; Mismatches 41; Indels 2; Gaps 2;  
QY 314 TGTGAGGACAAAGAAATTCGCTCAACAGACAGCTAGCTCCCTGGTGTGTTCAACT 373  
Db 119057 TGAGAGACAAAGGAGGAGTTGTTCAACATGACTCA-ATAGCTGATGAGTTCAAT 118999  
QY 374 TTGAGAGTTTGACCAACAGACACTTTATTTTGAATATTTAAACAGACAACTTTGGG 433  
Db 118998 TTGAAGTCTGACCTCAAGATTTTATTTCAAGCATATTTAAGCACAGAGAA-ATTGAA 118940  
QY 434 AAAAGTTTCTTATGAAAAATTTATCAAT 463  
Db 118939 AAGAACTTTGTAATATGATTAATAGTTAAT 118910

RESULT 12  
US-10-367-094-122  
; Sequence 122, Application US/10367094  
; Publication No. US20040170982A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc Malandro  
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
; FILE REFERENCE: 529452001500  
; CURRENT APPLICATION NUMBER: US/10/367,094  
; CURRENT FILING DATE: 2003-02-14  
; NUMBER OF SEQ ID NOS: 203  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 122  
; LENGTH: 421609  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(421609)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-367-094-122

Query Match 12.0%; Score 60; DB 18; Length 421609;  
Best Local Similarity 67.8%; Pred. No. 1.4e-05;  
Matches 99; Conservative 0; Mismatches 45; Indels 2; Gaps 1;  
QY 318 AGGCAAGAGAAATTCGCTCAACAGCTAGCTCCCTGCTGTTGTTCAACTTTGA 377  
Db 240077 AGGCAAGAGAAATTTAGCTCAACAGCTCAGTACCTG--TTCAATGTTGTTCA 240134  
QY 378 GAGTTTGACCAACAAGCACTTATTTTGAATATTTAAACAGACACACTTTGGAAA 437  
Db 240135 ACTTCTGACCTGAGCAGTTTATTTAGCCATTTTAAACAGACAGCAAGTTTGGAAAC 240194  
QY 438 AGTTTCTTATGAAATTTATCAAT 463  
Db 240195 AGTTTCTGATGCTATTTAATCAAT 240220

RESULT 13  
US-09-997-722-199  
; Sequence 199, Application US/09997722  
; Publication No. US20040072154A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71171/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/09/997,722  
; CURRENT FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 301  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 199  
LENGTH: 96599  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4189)..(4208)  
OTHER INFORMATION: "n" at positions 4189 through 4208 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (9652)..(10001)  
OTHER INFORMATION: "n" at positions 9652 through 10001 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (27805)..(27824)  
OTHER INFORMATION: "n" at positions 27805 through 27824 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (54062)..(54139)  
OTHER INFORMATION: "n" at positions 54062 through 54139 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (88386)..(88405)  
OTHER INFORMATION: "n" at positions 88386 through 88405 can be any base.  
US-09-997-722-199

Query Match 11.7%; Score 58.6; DB 11; Length 96599;  
Best Local Similarity 77.1%; Pred. No. 1.7e-05;  
Matches 84; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

Qy 387 CACAGACCTTATTTTGGACATTTTAACAGAGCACACTTGGAAAAAGTTTCTT 446  
Db 46014 CTCGAACGTTTCATTTTGAACATTTTAACAGAGCACA-TTTTGGAAAAAGTTTCTG 46072  
Qy 447 ATGAATAATTATCACAAATTAAGCTTAAGCATGACTTAAATGCTT 495  
Db 46073 GTGACAGTTATCACAGTAAGCCTTAAGGTGACTTAAATGAACTCTT 46121

RESULT 14  
US-10-085-117-235  
Sequence 235, Application US/10085117  
Publication No. US2003023334A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
FILE REFERENCE: 529452000121  
CURRENT APPLICATION NUMBER: US/10/085,117  
PRIOR FILING DATE: 2002-02-27  
PRIOR APPLICATION NUMBER: US 09/798,586  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 235  
LENGTH: 33488  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: variation  
LOCATION: (1)....(33488)  
OTHER INFORMATION: n = any nucleotide  
US-10-085-117-235

Query Match 11.6%; Score 58; DB 17; Length 33488;  
Best Local Similarity 66.1%; Pred. No. 1.5e-05;  
Matches 115; Conservative 0; Mismatches 55; Indels 4; Gaps 2;

Qy 293 GAGTGGACAGGACAGCAATGTGAGGCAGGAAGAAATTTGGCTCAACACAGCT--T 349

Db 5303 GATGGCCACAGACCAATCCAGAGGGGAAACAAAAGTGTCATTTTAACCAACTAAGT 5362  
Qy 350 AGCTCCCTGATGTGTGTTTAACTTTGAGAGTTTGAACACAGACCTTATTTTGGACAT 409  
Db 5363 AAGTACCGAGTATTTGTTTCAACTTTGAGAGTCTGACCCCAAGTATTTATTTAGGCTT 5422  
Qy 410 ATTTAAACAGACACAATTTTGGAAAAAGTTTCTTATGAATAATTAACAAT 463  
Db 5423 ATTTAAGCACAGACACAC-TTGACAGACATTTCTGTATTAATTAATTACTAGT 5475

RESULT 15  
US-10-087-192-1237  
Sequence 1237, Application US/10087192  
Publication No. US20020182586A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
PRIOR FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1237  
LENGTH: 75101  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)....(75101)  
OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-1237

Query Match 11.6%; Score 58; DB 13; Length 75101;  
Best Local Similarity 57.3%; Pred. No. 2.3e-05;  
Matches 126; Conservative 0; Mismatches 90; Indels 4; Gaps 1;

Qy 219 TATATTGCAATTTATGAGACAGCTAAATGTACTCGGGTATGATGACTCCAGTGGGATA 278  
Db 35225 TTTTATCCAAAATTCAGAAAGCAAAATGAGGGCTAAGATGGAACTGCTCAGTGTGGTT 35284  
Qy 279 TCTGCAAGTCCATGAGTGGGAGAGGACAGCCATGTGAGCAAGAAAGAAATTCGTGCT 338  
Db 35285 CCGCAAGCTTCAAGGGCTGGAAAGAAACAACTAGTGAAGTGAAGAAAGCTTGTCTT 35344  
Qy 339 CACACAGCTTAGCTCCCTGCTGTGTGTTCAAACTTTGAGAGTTTGAACCAAGCACTT 398  
Db 35345 CAGCCGACCTTAGCA----GGAGCCGGCCACACTTTGAAGTCTGACACTGGGTGTTT 35400  
Qy 399 ATTTTGAATATTTAAACAGACCAACTTTGGAAAAA 438  
Db 35401 ATTTTGAATATTTAAACAGACCAACTTTTGAAGAAAA 35440

Search completed: February 10, 2005, 18:16:57  
Job time : 372.063 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33 ; Search time 2171.78 Seconds  
(without alignments)  
8780.927 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_5000\_5500

Perfect score: 501  
Sequence: 1 ggtcagccatgcttactt.....acattaaatgcttgcga 501

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gsa1: \*  
9: gb\_gsa2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	24.8	768	9	CR142741 Reverse s
2	118.4	23.6	732	8	A2900293 RPT-24-2
3	99.2	19.8	470	8	A2560675 RPT-23-2
4	93.6	18.7	738	9	AG602339 Mus muscu
5	90.8	18.1	532	8	A2413362 RPT-23-1
6	85.6	17.1	750	8	A2555307 RPT-23-1
7	84.8	16.9	409	6	CB770238 AGNNDC:S
8	83.4	16.6	500	8	B2189040 CH230-378
9	82.8	16.5	812	8	B2257122 CH230-334
10	82.6	16.5	871	9	CR058176 Forward s
11	82.6	16.5	950	9	CR198109 Forward s
12	82	16.4	714	9	AG489174 Mus muscu
13	81.4	16.2	455	8	A2776986 2M0011N06
14	81	16.0	824	9	CR165051 Forward s
15	79.8	15.9	1506	3	AK089860 Mus muscu
16	79.4	15.8	780	9	AG548396 Mus muscu
17	79.4	15.6	478	1	CR003773 Forward s
18	77.8	15.4	478	1	A1464155 vc38h12.x
19	77.2	15.3	644	8	A2497328 IM0334F07
20	76.6	15.3	903	8	CR109486 Forward s
21	76.2	15.2	639	8	A2653636 IM0527H20
22	76.2	15.2	2135	3	AK085732 Mus muscu
23	76	15.2	445	2	AW495957 up45a02.y
24	75.8	15.1	730	9	AG433222 Mus muscu

C 25	75.6	15.1	668	9	CR180254 Reverse s
C 26	75.6	15.1	779	9	CR249056 Reverse s
C 27	75.6	15.1	784	9	CR240016 Reverse s
C 28	75.4	15.0	562	8	A2744586 RPT-24-1
C 29	75.2	15.0	584	8	A2090387 RPT-23-4
C 30	75	15.0	452	9	BX977127 Forward s
C 31	74.8	14.9	1282	9	AG321227 Mus muscu
C 32	73.8	14.7	752	9	AG324894 Mus muscu
C 33	73.4	14.7	612	8	AO997450 RPT-23-3
C 34	73.4	14.7	672	8	CR145845 Forward s
C 35	73	14.6	566	8	A2028328 RPT-23-2
C 36	72.8	14.5	628	8	B2126387 CT230-393
C 37	72.6	14.5	628	8	A2831052 2M0110B10
C 38	71.2	14.2	569	1	A1430674 mc56h07.y
C 39	70.6	14.1	471	8	A2592768 1M0403120
C 40	70.4	14.1	714	8	BH049545 RPT-24-3
C 41	70	14.0	769	6	CB325720 UT-R-DZO-
C 42	69.8	13.9	739	9	AG444044 Mus muscu
C 43	69.8	13.9	1682	3	AK082500 Mus muscu
C 44	69.6	13.9	645	2	BB624223 BB624223
C 45	69.6	13.9	2708	3	AK033143 Mus muscu

#### ALIGNMENTS

RESULT 1  
CR142741 768 bp DNA linear GSS 06-JUL-2004  
LOCUS  
DEFINITION Reverse strand read from insert in 3'HPRT insertion targeting and chromosome engineering clone MHP42h15, genomic survey sequence.

ACCESSION CR142741 GI:49890618  
VERSION CR142741.1  
KEYWORDS GSS; genome survey sequence; MICEP.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
1 (bases 1 to 768)  
Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,  
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,  
Rogers,J. and Bradley,A.  
TITLE  
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgehire,  
CB10 1SA, UK. http://www.sanger.ac.uk/MICEP

FEATURES  
source  
1..768  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="MHP42h15"  
/clone\_1ib="MHP"

#### ORIGIN

Query Match 24.8%; Score 124; DB 9; Length 768;  
Best Local Similarity 73.7%; Pred. No. 4.7e-23; Indels 1; Gaps 1;  
Matches 171; Conservative 0; Mismatches 60;  
270 GTGGGATATCTGCAAGTGCATGATGCGAGGACACCAATGTAGGCAAGAAGA 329  
|||||  
381 GTGGGGTCCCTTCAACATTAATGCTGCAAAAAAATGATGCTCATGAGCAAAAAA 440  
|||||  
330 ATTTCGCTGACACAGCTTACCTCTGCTGTTGTTCAACTTTAGAGTTGACAC 389  
|||||  
441 AATTCAGTCAACA-AATTAGTGAAGTCTGGGTTCAAACTTTAAGTGAACCC 499  
|||||  
390 AAGCACTTTTATTTGACATATTTAAACAGACCACTTTGGAAAAAGTTTCTATG 449  
|||||  
500 GAGTAGTTTATTTTGTATTAATTAACATTAAGTCAAGTCAAGTCAAGTCAAGT 559  
|||||  
450 AAATTTATCAATTAAGTTAAGGATGATGATTAATTAAGTCAAGTCAAGTCAAG 501  
|||||  
560 ACAAATTTTCAATTAAGTTAAGGATGATGATTAATTAAGTCAAGTCAAGTCAAG 611  
|||||

RESULT 2  
 AZ900293  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AZ900293 732 bp DNA linear GSS 05-MAR-2001  
 RPCI-24-212G17.TU RPCI-24 Mus musculus genomic clone  
 RPCI-24-212G17, genomic survey sequence.  
 AZ900293  
 AZ900293.1 GI:13219238  
 GSS.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 732)  
 Zhao, S., Niernan, W., Malek, J., Shatsman, S., Akintret, B., Levins, M.,  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P., and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 Unpublished (1999)  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 212 row: G column: 17  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..732  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-212G17"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_id="RPCI-24"  
 /note="Vector: pTARBAcl; Site 1: BamHI; Site 2: BamHI;  
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
 library was cloned in the pTARBAcl cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."

ORIGIN  
 Query Match 23.6%; Score 118.4; DB 8; Length 732;  
 Best Local Similarity 75.0%; Pred. No. 1,7e-21;  
 Matches 174; Conservative 0; Mismatches 56; Indels 2; Gaps 2;

270 GTGGGGAATATCTGCAAGTCCATGATGCGACAGGACACGACCAATGTGAGGCAAGAAGA 329  
 55 GTGGGGTCCCTCAAAACATTAATGATGCTGACAGATGATGCTCATGAGGCAAGAAAGA 114  
 330 ATTCTGGCTCAACACAGCTTACCTCCCTGGTGTGTTCAAACTTTAGAGTTGACAC 389  
 115 AATCAGTTCAACAGAT-TTAGTTAGTGTGCTGGGTTCAAACTTTAGAGTTGACAC 173  
 390 AAGCACTTATTTTGAATATTTAAACAGACCAACTTTGGAAAAAGTTTCTTATG 449  
 174 GAGTAGTTATTTTGA-TTCTTAAGCATGTACAACTTTGAAAAAGTTTCTTATG 232  
 450 AAAATTATCACAATAAAGCTTAAGGATGATTAATTAATGCTTTGGCA 501  
 233 ACAATTATCACAATAAAGTTTGAAGCATGACTTAAATACTCTTTGCA 284

RESULT 3  
 AZ560675/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AZ560675 470 bp DNA linear GSS 20-NOV-2000  
 RPCI-23-216E16.TU RPCI-23 Mus musculus genomic clone  
 RPCI-23-216E16, genomic survey sequence.  
 AZ560675  
 AZ560675.1 GI:11240495  
 GSS.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 470)  
 Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatsman, S.,  
 Akintret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de  
 Jong, P., and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 Other GSSs: RPCI-23-216E16.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
 or from Resea ch Genetics ([info@resgen.com](http://info@resgen.com)). BAC end page:  
[http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 216 row: E column: 16  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..470  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-216E16"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /clone\_id="RPCI-23"  
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:  
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBAC3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN  
 Query Match 19.8%; Score 99.2; DB 8; Length 470;  
 Best Local Similarity 69.7%; Pred. No. 3e-16;  
 Matches 163; Conservative 0; Mismatches 68; Indels 3; Gaps 2;

270 GTGGGGAATATCTGCAAGTCCATGATGCGACAGGACACGCAATGTGAGGCAAGAAGA 329  
 360 GTGAAGGCTCTCGCAACACTATGCAACAGAGAAAGTCACAGGGAGGCAAGGAA 301  
 330 ATTCTGGCTCAACACAGCTTACCTCCCTGGTGTGTTCAAACTTTAGAGTTGACAC 389  
 300 AATCTGTTCAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 241  
 390 AAGCACTTATTTTGAATATTTAAACAGACCAACTTTGGAAAAAGTTTCTTATG 449  
 240 GGGCAATTTGGTTTGCACACATTTAAGTTCAGTAC-AGTTTGGAAAAATATTTCTG 182  
 450 AAAATTATCACAATAAAGCTTAAGGATGATTAATTAATGCTTTGGCA 501  
 181 ACAATTATCACAATAAAGGCTTTAAGCATGCTTAACTCTTTGCA 128

AZ413362/c		532 bp	DNA	linear	GSS 03-OCT-2000
LOCUS					
DEFINITION	U0197G04F Mouse 10kb plasmid UUGCM library Mus musculus genomic clone UUGCM0197G04 F, genomic survey sequence.				
ACCESSION	AZ413362				
VERSION	AZ413362.1	GI:10537375			
KEYWORDS	CDS,				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 532)				
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacom,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingy,A., von Niederhausen,A. and Wright,D.,Weiss,R. Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000) Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddun@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0197 row: G column: 04 Seq primer: CGTTGTAAAGCAGCGCACAGT Class: plasmid ends High quality sequence stop: 532. Location/Qualifiers 1..532 /organism="Mus musculus" /mol_type="genomic DNA" /archiv="CS7BL/6J" /db_xref="taxon:10090" /clone="UUGCM0197G04" /bex="Male" /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-" /clone_lib="Mouse 10kb plasmid UUGCM library" /note=Vector: PWD42nv; Purified genomic DNA from M. musculus C57Bl/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource <a href="http://www.jax.org/resources/documents/dnares/">(http://www.jax.org/resources/documents/dnares/)</a> . The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 [gi 4732114 gb AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."				
ORIGIN					
Query Match	18.1%; Score 90.8; DB 8; Length 532; Best Local Similarity 66.7%; Pred. No. 6.5e-14;				
Matches 160; Conservative 0; Mismatches 77; Indels 3; Gaps 2;					
OY	262 GAATTCAGTGGGGAATTATTCGCAGAATGCCAATGACTGAGAGCAAGACCACCATGTGAGACG 321				
DB	513 GTCTTTTAGAGGGGGTTCCTACAAGTTCCTCTGTGTCAGAAAACAATTGCAAGAGCT 454				
OY	322 AAAGAAGAAATTCGGCTCAACACAGCTAGCTCCCTGGTGTGTTCAAACCTTGAAGAGT 381				

Db 453 AAGACAGAACCTGTCAGTACACTTGTGGCCTG--GCTCATTCAAACCTTGTGAGT 396

QY 382 TTGACCACAGCACTTTATTTTGGACATATTTAAACAGACAACTTTGGAAAAAGTT 441

Db 395 CTGAGCCCAATGATTTATTTTGGACATATTTAAATAGTACAACTTTGGG-AAAGTTT 337

QY 442 TTTCTATGAAAATATTCACATTAAGCTTAAGGCTGACTACATTAATAATGCTTTGGAA 501

Db 336 TCCCTGTGACATATTCACATTAAGCTTCGGGCAATGGCTGCAATTAATACTCTTGCAA 277

RESULT 6  
AZ555307 750 bp DNA linear GSS 20-NOV-2000  
LOCUS RPCI-23-175N14.TJ RPCI-23 Mus musculus genomic clone  
DEFINITION RPCI-23-175N14, genomic survey sequence.  
ACCESSION AZ555307  
VERSION AZ555307.1 GI:11235127  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 750)  
Zhao, S., Niernann, W., Feldblyum, T., Malek, J., Shatsman, S., Akimel, B., Levins, M., McGarr, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea.ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac/ends/mouse/bac\_end\_intro.html  
plate: 175 row: N column: 14  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..750  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-175N14"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: PBAC3.6; Site 1: EcoRI, Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the PBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN  
Query Match 17.1%; Score 85.6; DB 8; Length 750;  
Best Local Similarity 66.5%; Pred. No. 2e-12;  
Matches 157; Conservative 0; Mismatches 69; Indels 10; Gaps 2;

QY 270 GTGGGCAATCTGCAAGTGCATGAGCGACAGAGGACACCCATGTGAGCAAGAAGA 329

Db 253 GTGAGAGTCCCCCCCAACACTTAAGGCTTACTGGGAGCAAAACAGAAAGCAAGAAGA 312

QY 330 ATTCTGACTCAACACAGCTTAGCTCCTGGTGTGTTCAAACTTTGAGAGTTGACAC 389

Db 313 AATCTGCTCAGTGAAGACTTAGCTGGGCGG-----TTCAAACCTTCGGCGTTGACTCT 366

QY 390 AAGACATTATTTTGAACATTTTAAACA----GAGACAACTTTGGAAAAAAGTTTCT 445

Db 367 AAACAGTTTACTTTATTTATTTTGAACATTTTGAAGACAACTTGGGAAAAAGTTTCT 426

QY 446 TATGAATAATTCACATTAAGCTTAAGGCACTGACTACATTAATAATGCTTTGCAA 501

Db 427 GATGACATTAACAAATTAAGGCTTAAGGCACTGACTACATTAATAATGCTTTGCAA 482

RESULT 7  
CB770238 409 bp mRNA linear EST 16-MAY-2003  
LOCUS AMANNND:SRCP2-00001-E2-A srp2 (10233) Rattus norvegicus cDNA clone  
DEFINITION srp2-00001-e2 5', mRNA sequence.  
ACCESSION CB770238  
VERSION CB770238.1 GI:29858629  
KEYWORDS EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 409)  
Amgen EST Program.  
Amgen Rat EST Program  
Unpublished (2003)  
Contact: Dan Fitzpatrick  
Amgen, Inc  
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA  
Tel: 805 447-4881  
Plate: 00001 row: e column: 2.

FEATURES  
source Location/Qualifiers  
1..409  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="srp2-00001-e2"  
/release\_type="choroid plexus brain"  
/clone\_lib="srp2 (10233)"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; choroid plexus brain region"

ORIGIN  
Query Match 16.9%; Score 84.8; DB 6; Length 409;  
Best Local Similarity 71.1%; Pred. No. 2.8e-12;  
Matches 128; Conservative 0; Mismatches 47; Indels 5; Gaps 1;

QY 317 GAGGCAGAAGGAATTTGCTCAACACAGCTTAGCTCCCTGGTGTGTTGAAACTTGG 376

Db 43 GAGTTAATAGAAATCTTGTTGCAATTAAGCTTACT-----GGCCCGCTTCAGACCTCT 97

QY 377 AAGGTTTGACCAAGCACTTATTTTGGACATATTTAAACGAGACACAACTTTGGGAAA 436

Db 98 GGAGCTGACACCAAGCACTTATTTTGGACATATTTAAACCAAGTAAACCTTTGGGGA 157

QY 437 AAGTTTCTTAAGAAAATATTAACAATAAGCTTAAGGCACTGACTACATTAATAATGCTT 496

Db 158 AAGTTTCTTGTGATTAATATCAACAAAGCTCCAGGCTTGCTGCTTGAATCTCAT 217

RESULT 8  
B2189040/c 500 bp DNA linear GSS 11-OCT-2002  
LOCUS CH230-378B15.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone  
DEFINITION CH230-378B15, genomic survey sequence.  
ACCESSION B2189040  
VERSION B2189040.1 GI:22843505  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 500)  
 AUTHORS Zhao, S., Shetty, J., Shatman, S., Teegaye, G., Geer, K.,  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 TITLE Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,  
 Riggs, F., de Jong, P. and Fraser, C.M.  
 JOURNAL Rat BAC End Sequences from Library CHORI-230 MboI segment  
 COMMENT Unpublished (1999)  
 Other GSSs: CH230-378B15.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
 availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
 Clones may be purchased from BACPAC Resources  
 (<http://www.chori.org/bacpac/orering/information.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)  
 Plate: 378 row: B column: 15  
 Class: BAC ends.  
 Seq primer: Sp6  
 Location/Qualifiers  
 1..500  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SsNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-378B15"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_1lib="CHORI-230 Segment 2"  
 /note="Vector: PTARBAC1.3; Site\_1: MboI; Site\_2: MboI;  
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by  
 Pieter de Jong"

ORIGIN  
 Query Match 16.6%; Score 83.4; DB 8; Length 500;  
 Best Local Similarity 75.8%; Pred. No. 7.1e-12;  
 Matches 116; Conservative 0; Mismatches 36; Indels 1; Gaps 1;  
 QY 349 TAGCTCCCTGGTGGTGAACCTTGAAGTGAACAAGACCTTTATTTTGA 408  
 DB 283 TTGTTAGCTGGTTTGAAGTTCGCGAGTCTGACC-CAGCAGTTATTTTGA 225  
 QY 409 TATTAAACAGACAACTTTGGAAAAAGTTTCTTATGAAAAATATCAATTAAGC 468  
 DB 224 TCGTTAAACATAGCAACCTTGAATAAAAGTCTCTGTCCAATTACATGATAAGC 165  
 QY 469 TTAAGCATGACTACATTAATAATGCTTTGCAA 501  
 DB 164 TTAAGCATGACCCCATGAAATCTCTTGCAA 132

RESULT 9  
 B2257122 812 bp DNA linear GSS 15-OCT-2002  
 LOCUS CH230-334G8.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 DEFINITION CH230-334G8, genomic survey sequence.  
 ACCESSION B2257122  
 VERSION B2257122.1 GI:23963817  
 KEYWORDS GSS.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 812)  
 AUTHORS Zhao, S., Shetty, J., Shatman, S., Teegaye, G., Geer, K.,

REFERENCE 1 (bases 1 to 500)  
 AUTHORS Zhao, S., Shetty, J., Shatman, S., Teegaye, G., Geer, K.,  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 TITLE Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,  
 Riggs, F., de Jong, P. and Fraser, C.M.  
 JOURNAL Rat BAC End Sequences from Library CHORI-230 MboI segment  
 COMMENT Unpublished (1999)  
 Other GSSs: CH230-334G8.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
 availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
 Clones may be purchased from BACPAC Resources  
 (<http://www.chori.org/bacpac/orering/information.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)  
 Plate: 334 row: G column: 8  
 Class: BAC ends.  
 Seq primer: Sp6  
 Location/Qualifiers  
 1..812  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SsNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-334G8"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_1lib="CHORI-230 Segment 2"  
 /note="Vector: PTARBAC1.3; Site\_1: MboI; Site\_2: MboI;  
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by  
 Pieter de Jong"

ORIGIN  
 Query Match 16.5%; Score 82.8; DB 8; Length 812;  
 Best Local Similarity 67.5%; Pred. No. 1.2e-11;  
 Matches 131; Conservative 0; Mismatches 62; Indels 1; Gaps 1;  
 QY 270 GTGGGATATCTGCAAGTGCATGAGTGGAGGACCAAGTGTAGGCAAGCA 329  
 DB 67 GGGGTGTCTCTGAGGCACACAGTCAGACATGATCTGTGGAGGAAGAGT 126  
 QY 330 ATTCTGCTCAACAGCTTACCTCTGTTGGTTCAACTTGAAGATTGACAC 389  
 DB 127 AGTTTGCTTAAACA-AGACTATATAGCCAGTGTGTTCAAACTTGGGAGTCTGACCC 185  
 QY 390 AAGCACTTATTTTGCATATTTAAACAGACAACTTGGAAAAAGTTTCTTATG 449  
 DB 186 AAGTAGTTTATGTAGGCATATTCACACAAACAAATTTGGTAAAAATTTCTGGAG 245  
 QY 450 AAAATTATCACAAT 463  
 DB 246 ATATTGACTCAAT 259

RESULT 10  
 CR058176/c 871 bp DNA linear GSS 05-JUN-2004  
 LOCUS CR058176  
 DEFINITION Forward strand read from insert in 5'HPRT insertion targeting and  
 chromosome engineering clone MHPN417h21, genomic survey sequence.  
 ACCESSION CR058176  
 VERSION CR058176.1 GI:49791648  
 KEYWORDS GSS; genome survey sequence; MICE.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 871)  
 AUTHORS Adams, D.J., Biggs, P.J., Cox, A.V., Davies, R.M., van der Weyden, L.,  
 Jonkers, J., Smith, J., Plumb, R.W., Taylor, R.G., Nishijima, I., Yu, Y.,  
 Rogers, J. and Bradley, A.

TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. <http://www.sanger.ac.uk/MICER>  
FEATURES  
source  
1. 871  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="MHPN417h21"  
/clone\_lib="MHPN"  
ORIGIN  
Query Match 16.5%; Score 82.6; DB 9; Length 871;  
Best Local Similarity 61.7%; Pred. No. 1.4e-11;  
Matches 171; Conservative 0; Mismatches 94; Indels 12; Gaps 2;  
232 TGAGCAGCTAAATGTAATCTCGGGTACATGCTCCAGGTGGGATATCTGCAAGTGC 291  
DB 722 TCACACAGCTTACACTGACCTTACACCCCGATCTACTAGGGCTCTGCAATACCA 663  
QY 292 TGAGTGGCAGAGGAGCAGCCAAATGTGAGGCAAGAAATTCGTCTCAACAGCTTGA 351  
DB 662 TGCCTAGCTGATTTGTTATTCATGTGAGGCAAGAAATCCTGTTCAACAGACTCG 603  
QY 352 CTCCTGCTGTTGTTCAAACTTTGAGAGTTGACCAAGCAACTTATTTTGACATAT 411  
DB 602 AGGCC-GGTGTGTGTCACAACTTTGAGAGTCCAAATCTGGGTGTTATTTTAAAGCAC 544  
QY 412 TTAACAGAGCAACCTTTGGGAAAAAGTTTCTTATGAAAA-----TTATCAC 460  
DB 543 TTAGACAGCAGCAATTTGTAAGAAAAAGATTTCTGTGATTAAGTCAATCTGTGACAC 484  
QY 461 AATAAGCTTAAGCAGTACATCAATTAATGCTTT 497  
DB 483 AACAAATCATAGATATGATTAATCAATTAATCTTAT 447  
RESULT 11 950 bp DNA linear GSS 06-JUL-2004  
LOCUS CR199109/c  
DEFINITION Forward strand read from insert in 5'HPRT insertion targeting and  
chromosome engineering clone MHPN34g06, genomic survey sequence.  
ACCESSION CR199109  
VERSION CR199109.1 GI:49977958  
KEYWORDS GSS; genome survey sequence; MICR.  
SOURCE  
ORGANISM Mus musculus (house mouse)  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 950)  
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,  
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,T., Yu,Y.,  
Rogers,J. and Bradley,A.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. <http://www.sanger.ac.uk/MICR>  
FEATURES  
source  
1. 950  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="MHPN34g06"  
/clone\_lib="MHPN"  
ORIGIN  
Query Match 16.5%; Score 82.6; DB 9; Length 950;  
Best Local Similarity 61.7%; Pred. No. 1.4e-11;  
Matches 171; Conservative 0; Mismatches 94; Indels 12; Gaps 2;  
232 TGAGCAGCTAAATGTAATCTCGGGTACATGCTCCAGGTGGGATATCTGCAAGTGC 291  
DB 725 TCACACAGCTTACACTGACCTTACACCCCGATCTACTAGGGGTCTGCAATACCA 666

QY 292 TGAGTGGCAGAGGAGCAGCCAAATGTGAGGCAAGAAATTTCTGCTCAACAGCTTAG 351  
DB 665 TGCCTAGCTGATTTGTTATTCATGTGAGGCAAGAAATTCCTGTTCAACAGACTCAG 606  
QY 352 CTCCTGCTGTTGTTCAAACTTTGAGAGTTGACCAAGCAACTTATTTTGACATAT 411  
DB 605 AGGCC-GGTGTGTGTCACAACTTTGAGAGTCCAAATCTGGGTGTTATTTTAAAGCAC 547  
QY 412 TTAACAGAGCAACCTTTGGGAAAAAGTTTCTTATGAAAA-----TTATCAC 460  
DB 546 TTAAGCAGCAGCAATTTGTAAGAAAAAGATTTCTGTGATTAAGTCAATCTGTGACAC 487  
QY 461 AATAAGCTTAAGCAGTACATCAATTAATGCTTT 497  
DB 486 AACAAATCATAGATATGATTAATCAATTAATCTTAT 450  
RESULT 12 714 bp DNA linear GSS 04-JUN-2004  
LOCUS AG489174  
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-386C04.TJ, genomic survey  
sequence.  
ACCESSION AG489174  
VERSION AG489174.1 GI:48196404  
KEYWORDS GSS.  
SOURCE  
ORGANISM Mus musculus molossinus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
TITLE BAC end Sequences of Library MSMg01  
JOURNAL Unpublished  
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
2 (bases 1 to 714)  
TITLE Direct Submission  
JOURNAL Submitted (17-NOV-2003) Maehara Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Sushiro-cho,Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@gsc.riken.jp, URL:<http://hgp.gsc.riken.go.jp/>,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kunihisa Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyada, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : TJ  
LIBRARY : pBACe3.6  
Vector : EcoRI  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
FEATURES  
source  
1. 714  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-386C04.TJ"  
/sex="male"  
/tissue type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"  
ORIGIN  
Query Match 16.4%; Score 82; DB 9; Length 714;  
Best Local Similarity 69.2%; Pred. No. 1.9e-11;  
Matches 144; Conservative 0; Mismatches 55; Indels 9; Gaps 2;  
298 GCAGAGGAGCAGCCAAATGTGAGGCAAGAAATTTCTGCTCAACAGCTTAGCTCCCT 357  
DB 371 GCTGGAGACAAACAGAAAGACAGAAAGAAATCTGCTCAGCAGAACTTAGCT---- 426

358 GGTGTTGTTCAAACTTTGAGTTGACCAAGACCTTATTTTGCATATTTAAAC 417  
427 -GGGCGGGTTCAAACTTTGCGCTTGACTTAAAGATTTACTTATTTATTTTGGACA 485  
418 A-----GAGCACAATTGGGAAAAGTTTCTATGAGAAATTTCCAAATTAAGCTTAAG 473  
486 TATTTAAAGCAACTCTGGGAAAAGCTTCTGATATACATATCAATTAAGGTTAAG 545  
474 GCATGACTACATTAATAATGCTTTGGCA 501  
546 GCATGCTACTATGAAACCTCTTACTA 573

RESULT 13  
A2776986 455 bp DNA linear GSS 16-FEB-2001  
LOCUS 2M0011N06F Mouse 10kb plasmid UGCG1M library Mus musculus genomic  
DEFINITION clone UGCG2M0011N06 F, genomic survey sequence.  
ACCESSION A2776986  
VERSION A2776986.1 GI:12905133  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 455)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Redlhauser, A., and Wright, D., Weisse, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0011 row: N column: 06  
Seq primer: CGTTGTAAGACGACGCCACAT  
Class: plasmid ends  
High quality sequence stop: 455.  
Location/Qualifiers  
1..455  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCG2M0011N06"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_id="Mouse 10kb plasmid UGCG1M library"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to

adapored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN  
Query Match 16.2%; Score 81.4; DB 8; Length 455;  
Best Local Similarity 68.7%; Pred. No. 2.5e-11;  
Matches 112; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

271 TGGGATATCTGCAAGTGCATGAGTGCGAGGAGCAGCCAAATGAGGCAAGAGAA 330  
45 TGGGGGCTCTGTAAGAACACCAATTCCTGACAGATATTAATCTGACAGAGTAAGAGCA 104  
331 TTCTGGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAACTTTGAGAGTTGACACA 390  
105 ATCTGCTTATCAACATGATATCTTCTGGTGTGTTCAACTTTGATGCTGACCCCA 164  
391 AGCACTTATTTTGAATATTTAAACAGAGCAACCTTTGGG 433  
165 GGCATTCATTTCTTGACATATTTTACAGCGAGTGCAATCTGTG 207

RESULT 14  
CRI65051 824 bp DNA linear GSS 06-JUN-2004  
LOCUS CRI65051  
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and chromosome engineering clone MHPJ34b03, genomic survey sequence.  
ACCESSION CRI65051  
VERSION CRI65051.1 GI:49943900  
KEYWORDS GSS; genome survey sequence; MICR.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 824)  
AUTHORS Adams, D.J., Biggs, P.J., Cox, A.V., Davies, R.M., van der Weyden, L., Jonkers, J., Smith, J., Plumb, R.W., Taylor, R.G., Nishijima, I., Yu, Y., Rogers, J., and Bradley, A.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICR  
Location/Qualifiers  
1..824  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="MHPJ34b03"  
/clone\_1lb="MHP"

ORIGIN  
Query Match 16.2%; Score 81; DB 9; Length 824;  
Best Local Similarity 70.3%; Pred. No. 3.8e-11;  
Matches 137; Conservative 0; Mismatches 55; Indels 3; Gaps 2;

270 GTGGGATATCTGCAAGTGCATGAGTGCGAGGAGCAGCCAAATGAGGCAAGAGAA 329  
86 GCGAGGCTCTGTAAGAACACCAATTCCTGACAGATATTAATCTGACAGAGTAAGAGCA 145  
330 ATTCTGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAACTTTGAGAGTTGACAC 389  
146 AATCTGCTCAACATGATATCTTAGTACCAATGTTGATCAAACTTTAGAGTGAACCA 205  
390 AAGCACTTATTTT-GACATATTTAAACAGAGCAACCTTTGGGAAAAGTTTCTTAT 448  
206 GGGGCTTATTTTATGACATATTTAAGACATCACAATTTGG--AGAAATTTCTGTG 263  
449 GAAATATTCACAAAT 463  
264 GACATATCACTCAAT 278

RESULT 15  
AK089860/c

LOCUS AK089860 1506 bp mRNA linear HTC 03-APR-2004  
DEFINITION Mus musculus activated spleen cDNA, RIKEN full-length enriched library, clone: F830034A08 product: unclassified, full insert sequence.  
ACCESSION AK089860  
VERSION AK089860.1 GI:26105606  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PubMed 10349636  
REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20493374  
PubMed 11042159  
REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Ogawa, K., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PubMed 11076861  
REFERENCE  
AUTHORS  
TITLE The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
JOURNAL Nature 409, 685-690 (2001)  
MEDLINE  
PubMed  
REFERENCE  
AUTHORS  
TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
JOURNAL Nature 420, 563-573 (2002)  
MEDLINE  
PubMed  
REFERENCE  
AUTHORS  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
MEDLINE  
PubMed  
REFERENCE  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Adenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/.

## FEATURES

## source

1..1506  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NOD"  
/db\_xref="PANTOM DB:F830034A08"  
/db\_xref="taxon:10090"  
/clone="F830034A08"  
/tissue\_type="activated spleen"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
1..1506  
misc\_feature  
/note="unclassified"

## ORIGIN

Query Match 16.0%; Score 80; DB 3; Length 1506;  
Best Local Similarity 75.0%; Pred. No. 8.5e-11;  
Matches 126; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

QY 334 TGGCTCAACACAGCTTACCTCCGTGTGGTTCCTGAAGAGTTGACCAACAGC 393  
DB 805 TGGAAACAAACAACTTATGTTACTCTG-ACGTGTCACACTCTAGAACTACCTATAC 747  
QY 394 ACTTATTTTGGACATATTTAAACAGACACAACTTTGGGAAAAAGTTCTTATGAAA 453  
DB 746 AGTGA-TTTTCACATATATATAACACAGTGCACAACTTGGGAAAAAGTTCTTGTGATGA 688  
QY 454 TTATCAATTAAGCTTTAAGGATGATCATCTTAAATAGCCTTTGCAA 501  
DB 687 TTATTACAAACAAAGCTCAAGGCATTAATGAAAGTTCCATTCCAA 640

Search completed: February 10, 2005, 17:33:02  
Job time : 2176.78 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 21:17:58 ; Search time 2413.78 Seconds  
(without alignments)  
10057.285 Million cell updates/sec

Title: US-09-926-375B-7\_COPY\_11500\_12000

Perfect score: 501  
Sequence: 1 aaccacacagacatctctc.....gactcatctggtctcttctt 501

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

GenEmb1:.\*  
1: gb\_Da:.\*  
2: gb\_hlg:.\*  
3: gb\_in:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pac:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vl:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	13962	10	MMPSPG X68699 M. musculus
2	501	100.0	17732	6	AX042378 Sequence
3	501	100.0	20623	6	AX042372 Sequence
4	501	100.0	172535	10	AL732466 Mus muscu
5	501	100.0	214345	2	AC107726 Mus muscu
6	433.8	86.6	1219	10	MMPSPG X04713 Mouse parot
7	344.2	68.7	236802	2	AC103277 Rattus no
8	299.8	59.8	165313	10	AC125117 Mus muscu
9	284	56.7	284	10	MUSPSPAL M26805 Mouse parot
10	278.4	55.6	236802	2	AC103277 Rattus no
11	238.4	45.6	1733	10	AF153354 Rattus no
12	175.6	35.0	165313	10	AC125117 Mus muscu
13	175.6	35.0	211446	10	AC147545 Mus muscu
14	165.2	33.0	1196	10	AF153355 Rattus no
15	52.4	10.5	161593	9	HSBA49G10 Human DNA
16	45	9.0	62861	2	AC137726 Homo sapi
17	40.2	8.0	134341	9	AC099794 Homo sapi
18	39	7.8	190412	2	CR396584 Dario rer
19	38.8	7.7	21078	4	AY197556 Sus scrofa

20	38.6	7.7	244170	2	AC130152 Rattus no
21	37.8	7.5	160039	9	AC094109 Homo sapi
22	37.8	7.5	160039	9	AC068323 Homo sapi
23	37.8	7.5	195748	10	AL671889 Mouse DNA
24	37.6	7.5	240340	2	AC095354 Rattus no
25	37.6	7.5	270468	2	AC094628 Rattus no
26	37.4	7.5	5519	3	DMMDG3 X95908 D.melanogaster
27	37.4	7.5	74534	3	AC004735 Drosophila
28	37.4	7.5	82912	3	AC004574 Drosophila
29	37.4	7.5	94741	3	AE003607 Drosophila
30	37.4	7.5	159479	3	AC009735 Drosophila
31	37.4	7.5	164712	3	AC007589 Drosophila
32	37.4	7.5	166935	3	AC001920 Drosophila
33	37.4	7.5	169913	3	AC105898 Drosophila
34	37.4	7.5	170675	3	AC008139 Drosophila
35	37.4	7.5	170914	3	AC010843 Drosophila
36	37.4	7.5	179283	3	AC009383 Drosophila
37	37.4	7.5	179396	3	AC022351 Drosophila
38	37.4	7.5	184272	3	AC009211 Drosophila
39	37.4	7.5	185918	3	AC009839 Drosophila
40	37.4	7.5	190801	3	AC009212 Drosophila
41	37.4	7.5	191907	3	AC093157 Drosophila
42	37.4	7.5	203924	2	AC012159 Drosophila
43	37.4	7.5	223505	3	AE003686 Drosophila
44	37.4	7.5	247955	3	AE003698 Drosophila
45	37.4	7.5	296756	3	AE003492 Drosophila

## ALIGNMENTS

RESULT 1	
MMPSPG	
LOCUS	MMPSPG 13962 bp DNA linear ROD 29-OCT-1999
DEFINITION	M. musculus Psp gene for parotid secretory protein.
ACCESSION	X68699
VERSION	X68699.1 GI:53809
KEYWORDS	parotid secretory protein, mouse
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	Stendesen, P., Laurens, J., Krogh-Pedersen, H. and Hjorth, J. P.
TITLE	Novel salivary gland specific binding elements located in the PSP proximal enhancer core
JOURNAL	Nucleic Acids Res. 26 (11), 2761-2770 (1998)
MEDLINE	98256451
PUBMED	9592166
REFERENCE	2 (bases 1 to 13962)
AUTHORS	Mikkelsen, T. R.
TITLE	Direct Submissions
JOURNAL	Submitted (07-OCT-1992) T.R. Mikkelsen, Department of Molecular Biology, University of Aarhus, CF Mollers Alle 130, 8000 Aarhus, DENMARK
FEATURES	
source	Location/Qualifiers 1..13962 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C3H/As" /db_xref="taxon:10090" /chromosome="2" /map="Estimate: 69 cm from centromere" /clone="Lambda YP1, Lambda YP3, Lambda YP7" /clone_idb="Lambda-PHAGE (Lambda L47.1)" /germline /note="Allele: b" 1..1578 /evidence=experimental 4632..12433 /gene="Psp" 4632..4678 /gene="Psp" exon

```
/product="parotid secretory protein"
/Note="exon a"
/number=1
/evidence=experimental
5480..5608
/gene="Psp"
/product="parotid secretory protein"
/Note="exon b"
/number=2
6133..6274
/gene="Psp"
/product="parotid secretory protein"
/Note="exon c"
/number=3
6835..6942
/gene="Psp"
/product="parotid secretory protein"
/Note="exon d"
/number=4
7716..7868
/gene="Psp"
/product="parotid secretory protein"
/Note="exon e"
/number=5
10043..10124
/gene="Psp"
/product="parotid secretory protein"
/Note="exon f"
/number=6
10779..10842
/gene="Psp"
/product="parotid secretory protein"
/Note="exon g"
/number=7
11287..11361
/gene="Psp"
/product="parotid secretory protein"
/Note="exon h"
/number=8
12262..12433
/gene="Psp"
/product="parotid secretory protein"
/Note="exon i"
/number=9

ORIGIN
Query Match 100.0%; Score 501; DB 10; Length 13962;
Best Local Similarity 100.0%; Pred. No. 4.8e-145;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCAGAGACATTTCTTCATTGATTCATCTTGATCTTCTTTCTAGACATATCTGAA 60
DB 4354 AACCCAGAGACATTTCTTCATTGATTCATCTTGATCTTCTTTCTAGACATATCTGAA 4413
QY 61 CTGACCCCTGTTGGCAGCTCAAGTTTGACAATATGGCTTTGAACTTGACAAGT 120
DB 4414 CTGACCCCTGTTGGCAGCTCAAGTTTGACAATATGGCTTTGAACTTGACAAGT 4473
QY 121 CCATCACTGTCAACCAAGATCTGGGTGAAGTATCTTGGCTAACTT 180
DB 4474 CCATCACTGTCAACCAAGATCTGGGTGAAGTATCTTGGCTAACTT 4533
QY 181 AGATATTTCTTTGAGATCTTTTGAACATCCAGAAATAGGGCTTGATCTCATCTG 240
DB 4534 AGATATTTCTTTGAGATCTTTTGAACATCCAGAAATAGGGCTTGATCTCATCTG 4593
QY 241 GGAACCAATATTAAGTCAACCTTGAATCCAGAGATCTGACAGAAACAAGATCTCT 300
DB 4594 GGAACCAATATTAAGTCAACCTTGAATCCAGAGATCTGACAGAAACAAGATCTCT 4653
QY 301 CTGCTGTGATCTCTTCTTCAAGAGAGTGAAGTGAATCACTAACTGAATCCATCT 360
DB 4654 CTGCTGTGATCTCTTCTTCAAGAGAGTGAAGTGAATCACTAACTGAATCCATCT 4713
```

```
QY 361 GAGATCCACAGAGAGGCTTTGGGCAAGAGCAGAGGAGGCACTGTTGTGTGTA 420
DB 4714 GAGATCCACAGAGAGGCTTTGGGCAAGAGCAGAGGAGGCACTGTTGTGTGTA 4773
QY 421 AGTTTGAATCTTAACAAATTTGAAGACATAGATGATCTTGTGACATTAACAACCT 480
DB 4774 AGTTTGAATCTTAACAAATTTGAAGACATAGATGATCTTGTGACATTAACAACCT 4833
QY 481 AGACTCATGTGGGTCTGTTT 501
DB 4834 AGACTCATGTGGGTCTGTTT 4854

RESULT 2
AX042378 17732 bp DNA linear PAT 23-NOV-2000
LOCUS AX042378 Sequence 7 from Patent WO0064247.
DEFINITION AX042378
ACCESSION AX042378
VERSION AX042378.1 GI:11340996
KEYWORDS
SOURCE
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 Foreberg,C.W., Golovan,S. and Phillips,J.P.
AUTHORS Transgenic animals expressing salivary proteins
TITLE Patent: WO 0064247-A 7 02-NOV-2000;
JOURNAL UNIVERSITY OF GUELPH (CA)
FEATURES
source
location/Qualifiers
1..17732
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/Note="lamaz/APPA transgene"

ORIGIN
Query Match 100.0%; Score 501; DB 6; Length 17732;
Best Local Similarity 100.0%; Pred. No. 4.7e-145;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCAGAGACATTTCTTCATTGATTCATCTTGATCTTCTTTCTAGACATATCTGAA 60
DB 11500 AACCCAGAGACATTTCTTCATTGATTCATCTTGATCTTCTTTCTAGACATATCTGAA 11559
QY 61 CTGACCCCTGTTGGCAGCTCAAGTTTGACAATATGGCTTTGAACTTGACAAGT 120
DB 11560 CTGACCCCTGTTGGCAGCTCAAGTTTGACAATATGGCTTTGAACTTGACAAGT 11619
QY 121 CCATCACTGTCAACCAAGATCTGGGTGAAGTATCTTGGCTAACTT 180
DB 11620 CCATCACTGTCAACCAAGATCTGGGTGAAGTATCTTGGCTAACTT 11679
QY 181 AGATATTTCTTTGAGATCTTTTGAACATCCAGAAATAGGGCTTGATCTCATCTG 240
DB 11680 AGATATTTCTTTGAGATCTTTTGAACATCCAGAAATAGGGCTTGATCTCATCTG 11739
QY 241 GGAACCAATATTAAGTCAACCTTGAATCCAGAGATCTGACAGAAACAAGATCTCT 300
DB 11740 GGAACCAATATTAAGTCAACCTTGAATCCAGAGATCTGACAGAAACAAGATCTCT 11799
QY 301 CTGCTGTGATCTCTTCTTCAAGAGTGAAGTGAATCACTAACTGAATCCATCT 360
DB 11800 CTGCTGTGATCTCTTCTTCAAGAGTGAAGTGAATCACTAACTGAATCCATCT 11859
QY 361 GAGAGTCAACAGAGGCTTTGGGCAAGAGCAGAGGAGGCACTGTTGTGTGTA 420
DB 11860 GAGAGTCAACAGAGGCTTTGGGCAAGAGCAGAGGAGGCACTGTTGTGTGTA 11919
QY 421 AGTTTGAATCTTAACAAATTTGAAGACATAGATGATCTTGTGACATTAACAACCT 480
DB 11920 AGTTTGAATCTTAACAAATTTGAAGACATAGATGATCTTGTGACATTAACAACCT 11979
```

Fr1 Feb 11 08:00:43 2005

RESULT 3	AX042372	20623 bp	DNA	linear	PAT 23-NOV-2000
LOCUS	AX042372				
DEFINITION	Sequence 1 from Patent WO0064247.				
ACCESSION	AX042372				
VERSION	AX042372.1	GI:11340990			
KEYWORDS					
SOURCE	synthetic construct				
ORGANISM	synthetic construct				
REFERENCE	1	other sequences; artificial sequences.			
AUTHORS	1	Forsberg, C.W., Golovan, S. and Phillips, J.P.			
TITLE		Transgenic animals expressing salivary proteins			
JOURNAL		Patent: WO 0064247-A 1 02-NOV-2000;			
FEATURES	UNIVERSITY OF GUELPH (CA)				
source	Location/Qualifiers				
	1..20623				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
	/note="Lama2/Alpa plasmid"				
ORIGIN					
Query Match	100.0%; Score 501; DB 6; Length 20623;				
Best Local Similarity	100.0%; Pred. No. 4,7e-145;				
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 AACCAACAGAACATCTCCATGATTCATCTGATCTCTCTCTTCTAGCAATACTGAA	60			
DB	11500 AACCAACAGAACATCTCTTCATGATTCATCTGATCTCTCTCTCTTCTAGCAATACTGAA	11559			
QY	61 CTGACCCCTTGTGGCAGTCTCAAGTTTGACAACATATGGGCTTTGAACTTGGCAAAAGT	120			
DB	11560 CTGACCCCTTGTGGCAGTCTCAAGTTTGACAACATATGGGCTTTGAACTTGGCAAAAGT	11619			
QY	121 CCATCAGTGTGACCCAGACATCTGGGAGACCTTGGGTTGGAATATCTGGCTAACCTT	180			
DB	11620 CCATCAGTGTGACCCAGACATCTGGGAGACCTTGGGTTGGAATATCTGGCTAACCTT	11679			
QY	181 AGATATTTCTTTGGAGATCTTTAGAACATCCAGAAATAGGCTTGATTCATCTCG	240			
DB	11680 AGATATTTCTTTGGAGATCTTTAGAACATCCAGAAATAGGCTTGATTCATCTCG	11739			
QY	241 GGACCAACATATTAAGTCAACCTTAAGATCCAGAGATCTGACAGAAACAAAGATCTCT	300			
DB	11740 GGACCAACATATTAAGTCAACCTTAAGATCCAGAGATCTGACAGAAACAAAGATCTCT	11799			
QY	301 CTCGTGTCATCTCTCTCAAGAGAGTGAAGTCACTAACTGAGTCCCATCT	360			
DB	11800 CTCGTGTCATCTCTCTCAAGAGAGTGAAGTCACTAACTGAGTCCCATCT	11859			
QY	361 GAGAGTCCAGAGAGGCTTTGGGCAAGACAGAGGGAAGGCACTGTTTGTGGTAA	420			
DB	11860 GAGAGTCCAGAGAGGCTTTGGGCAAGACAGAGGGAAGGCACTGTTTGTGGTAA	11919			
QY	421 AGTTTGAAGTCTTAACAATTTGAAGACATAGATGACATTTGTGTGACATTAACAACCT	480			
DB	11920 AGTTTGAAGTCTTAACAATTTGAAGACATAGATGACATTTGTGTGACATTAACAACCT	11979			
QY	481 AGACTCATGTGGTCTGTCTT	501			
DB	11980 AGACTCATGTGGTCTGTCTT	12000			
RESULT 4					
LOCUS	AL732466	172515 bp	DNA	linear	ROD 24-OCT-2002
DEFINITION	Mouse DNA sequence from clone RP23-149G14 on chromosome 2, complete				

```

QY      241 GGACCACAATATTAAGTACCCCTTAGAATCCAGAGATGCTGCAGAGAAACAGATCTCT 300
DB      30863 GGACCACAATATTAAGTACCCCTTAGAATCCAGAGATGCTGCAGAGAAACAGATCTCT 30922
QY      301 CTCGGTGATCTCTTCAAGCAGTAGTAGTGACTGCACCTAAAGTAGTCCCATCT 360
DB      30923 CTCGGTGATCTCTTCAAGCAGTAGTAGTGACTGCACCTAAAGTAGTCCCATCT 30982
QY      361 GAGAGTCCACAGAGAGCTTTGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB      30983 GAGAGTCCACAGAGAGCTTTGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 31042
QY      421 AGTTTGAATCTTAACAAATTTGAAGACATGATGACTGTGTGCAGCTTAACAAACACT 480
DB      31043 AGTTTGAATCTTAACAAATTTGAAGACATGATGACTGTGTGCAGCTTAACAAACACT 31102
QY      481 AGACTGATGGGGTTCTGTTT 501
DB      31103 AGACTGATGGGGTTCTGTTT 31123

RESULT 5
AC107726 214345 bp DNA linear HTG 03-MAR-2003
LOCUS Mus musculus clone RP24-85E12, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
AC107726 GI:28631318
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren, B., Nussbaum, C. and Lander, E.
TITLE Mus musculus, clone RP24-85E12
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 214345)
REFERENCE
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
TITLE Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
JOURNAL Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
REFERENCE Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
AUTHORS Cook, P., DeArrelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
TITLE Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,
JOURNAL Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
REFERENCE Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
AUTHORS Kamat, A., Karatas, A., Kells, C., Labrecque, K., Lamazares, R.,
TITLE Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
JOURNAL Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M.,
REFERENCE McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
AUTHORS Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
TITLE Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
JOURNAL Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
REFERENCE Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, D.,
AUTHORS Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
TITLE Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
JOURNAL Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
REFERENCE Topham, K., Travers, M., Travis, N., Trigg, J., Vassiliev, H.,
AUTHORS Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
TITLE Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
JOURNAL Direct Submission
TITLE Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 214345)
AUTHORS Birren, B., Nussbaum, C., Lander, E., Aboueleil, A., Allen, N.,
TITLE Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
JOURNAL Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
REFERENCE Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrelano, K.,
AUTHORS Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
TITLE Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
JOURNAL Graham, L., Grand-Pierre, N., Hagos, B., Haploplan, D., Hays, B.,
REFERENCE Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
AUTHORS Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

```

```

Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Travis, N., Trigg, J., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.,
Direct Submission
TITLE Submitted (03-MAR-2003) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2003 this sequence version replaced gi:28173217.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 120103
Center clone name: 85 E 12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 212553 bases at least Q40
Consensus quality: 213324 bases at least Q30
Consensus quality: 213511 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 213645; sum-of-contigs
Quality coverage: 13.0 in Q20 bases; agarose-fp
Quality coverage: 11.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 64960: contig of 64960 bp in length
* 64961 65060: gap of 100 bp
* 65061 68275: contig of 3215 bp in length
* 68276 68376: gap of 100 bp
* 68376 74568: contig of 6193 bp in length
* 74569 74669: gap of 100 bp
* 74669 81021: contig of 6353 bp in length
* 81022 81121: gap of 100 bp
* 81122 91450: contig of 10329 bp in length
* 91451 91550: gap of 100 bp
* 91551 118992: contig of 27442 bp in length
* 118993 119092: gap of 100 bp
* 119093 165215: contig of 46123 bp in length
* 165216 165316: gap of 100 bp
* 165316 214345: contig of 49030 bp in length.
*
Location/Qualifiers
1. 214345
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-85E12"
/clone_1ib="RPCT-24 Male Mouse BAC"
1. 64960
/note="assembly_frgment"
/clone_end:SP6
/vector_side:left"
55061. 68275
misc_feature
misc_feature
misc_feature

```



```

RESULT 7
AC103277/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-235G1, *** SEQUENCING IN PROGRESS
*** 2 unordered pieces.
AC103277
AC103277.6 GI:30579004
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus
Rattus norvegicus (Norway rat)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 236802)
Muzny,D,Marie., Metzker,M,lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Bunay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Derano,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flaggs,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garner,M.,
Georgievski,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunatirup,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,T., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,S., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,
Jarsch,J., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolyet,A.,
Karpeth,S., Kelly,S., Kelly,S., Khan,Z., King,D., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshuwa,L., Loulsegad,H., Lozada,R.J., Lu,X., Ma,U.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Mijic,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Pawelczak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Pizarro,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Sma's,D.,
Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J.,
Steinle,M., Strong,R., Sutton,A., Swatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,B., Umani,K.,
Valas,R., Vera,V., Villaseña,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Woodson,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 236802)
Worley,K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE
3 (bases 1 to 236802)

```

```

AUTHORS
Rat Genome Sequencing Consortium.
TITLE
Direct Submission
JOURNAL
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:25086871.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJRO
Center clone name: CH230-235G1
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 226151 bases at least Q40
Consensus quality: 228208 bases at least Q20
Consensus quality: 228980 bases at least Q20
Estimated insert size: 234093; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 235405: contig of 235405 bp in length
* 235406 235505: gap of unknown length
* 235506 236802: contig of 1297 bp in length.
FEATURES
source
1..236802
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-235G1"
misc_feature
1..3075
/note="wgs_end_extension
clone_end:5p6"
misc_feature
complement(5172..5970)
/note="clone boundary
clone_end:5p6"
misc_feature
site:BCORI
end_sequence:B2095899"
228453..229719
/note="wgs_contig"
233558..233880
/note="clone_boundary
clone_end:T7
site:BCORI
end_sequence:B2095897"
ORIGIN
Query Match 68.7% Score 344.2; DB 2; Length 236802;
Best Local Similarity 86.8%; Pred. No. 3,9e-96;
Matches 440; Conservative 0; Mismatches 53; Indels 14; Gaps 5;

```



repeat_region	/rpt_family="MaLR"
	11977.12350
repeat_region	/rpt_family="MaLR"
	13628.13957
repeat_region	/rpt_family="ERVK"
	13960.14703
repeat_region	/rpt_family="ERVK"
	14700.14975
repeat_region	/rpt_family="ERVK"
	14976.15631
repeat_region	/rpt_family="ERVK"
	15607.15848
repeat_region	/rpt_family="ERVK"
	15893.16222
repeat_region	/rpt_family="ERVK"
	16911.17079
repeat_region	/rpt_family="B4"
	17169.17277
repeat_region	/rpt_family="Alu"
	19208.19300
repeat_region	/rpt_family="B4"
	23436.23667
repeat_region	/rpt_family="L1"
	28037.28164
repeat_region	/rpt_family="Alu"
	30431.31273
repeat_region	/rpt_family="ERVK"
	31532.31609
repeat_region	/rpt_family="B4"
	31833.31922
repeat_region	/rpt_family="Alu"
	31962.32291
repeat_region	/rpt_family="ERVK"
	32338.33367
repeat_region	/rpt_family="ERVK"
	37527.37856
repeat_region	/rpt_family="ERVK"
	37975.38424
repeat_region	/rpt_family="RMER12"
	38425.38774
repeat_region	/rpt_family="L1"
	39296.39555
repeat_region	/rpt_family="RMER12"
	40214.42773
repeat_region	/rpt_family="L1"
	42977.43025
repeat_region	/rpt_family="L1"
	43015.43516
repeat_region	/rpt_family="L1"
	45272.45464
repeat_region	/rpt_family="B4"
	51509.51627
repeat_region	/rpt_family="ERV1"
	51836.52252
repeat_region	/rpt_family="L1"
	52249.52290
repeat_region	/rpt_family="L1"
	52806.52922
unsure	/note="Sequence derived from one plasmid subclone."
	52864.52951
repeat_region	/rpt_family="L1"
	52953.53119
repeat_region	/rpt_family="B2"
	53160.53479
repeat_region	/rpt_family="L1"
	53530.53636
repeat_region	/rpt_family="L1"
	53637.54191
repeat_region	/rpt_family="ERVK"
	54515.54872
repeat_region	/rpt_family="L1"
	55568.55649
repeat_region	/rpt_family="B4"

```

repeat_region      56922. .57044  

                    /rpt_family="B4"  

repeat_region      57763. .57888  

                    /rpt_family="B2"  

repeat_region      58600. .58763  

                    /rpt_family="r1"  

repeat_region      58764. .58853  

                    /rpt_family="BRVK"  

repeat_region      64713. .64893  

                    /rpt_family="r1"  

repeat_region      64951. .65149  

                    /rpt_family="r1"  

repeat_region      65128. .65337  

                    /rpt_family="r1"  

repeat_region      66707. .66851  

                    /rpt_family="BRV1"  

repeat_region      66855. .66923  

                    /rpt_family="MALR"  

repeat_region      67272. .67468  

                    /rpt_family="B2"  

repeat_region      68460. .68569  

                    /rpt_family="B2"  

Query Match      59.8%; Score 299.8; DB 10; Length 165313;  

Best Local Similarity 83.5%; Pred. No. 2.9e-82;  

Matches 430; Conservative 0; Mismatches 62; Indels 23; Gaps 7;

```

QY		3	CCCAACGAACATTTCTTCATGATTTCATTTGGANCTTCTTTCTAGACAAATACGAAC	62
Dd		25000	CCCAACAACAATCTCTCATTCCTGCATCTTGCGGTCTTCTTGCAAGACAATACTGAATT	24941
QY		63	GACCCCTGTGGCAGCTCTCAAGTTTGAACAATAGGGGCTTTG-----AACTTGCA	114
Dd		24940	GA-CACATCTTGGCACGCTCAGGTTTCTCACTCAGGAATCAGTCTTGGCAGAATTCCT	24882
QY		115	CAAGTTCATCACTGTCAACCAGAATCTCGGTGAACCTTTGGGTGGAATACTTGGCT	174
Dd		24881	CTGCATCCACCACTGTAACCTTAAGTATCTGAGTGACCTTTGGGTGGAATCTTGGCT	24822
QY		175	AACCTTAGATATTTCTTTGGAGTATCTTTAAGACATCCAGAAATTAGGGCTGATTTCTC	234
Dd		24821	AACCTTAGATATTTCTTTGG-GTCTCTTTAAMAACTCCGGAAATGGACATCTGATTTCTC	24763
QY		235	ATCTCTGGACCAAAATATA-----AGTCACCTTAGAATCCAGAGATGCTGAGAGA	287
Dd		24762	ATCTCTGGACCAAAATATAAGGCCACAGTACCCCCAGAAATCCAGAGATCTGAGAGA	24703
QY		288	AACAAGATCTCTCTGCTGNGACCTCTTCTCAAAGCAGTGAATGATCTCCAATAAC	347
Dd		24702	AACAAGA----TCTGTGTGATCTCTTTCAAAGCAGTGAATGATCTCCAATAAC	24647
QY		348	TGAGTTCCCATCTGAGATGCCAAGAGGCTTTGGGCAAGAAGCAGAGGAGGCATCG	407
Dd		24646	TGAGTTCCCATCT-TTAGGCTTCAGAGGCTTTGGGCAAGAAGTAGAGGAGAGCGCTG	24588
QY		408	TTTGTGTGTGTAAGTTTGTGACT-CTAACAAATTTGAAGACATATGATGACATTTGTGTG	466
Dd		24587	TTTTGTGTGTAAGACGTTTGTGACTGTTAACAAATTTGMAACATATGATATTTGAGTCAG	24528
QY		467	ACTAACACAACCTAGACATCATGTGGGTTCTGTTT	501
Dd		24527	ACTAACACAAGAACCACTCATCATGGGTTCTGTTT	24493

RESULT 9					
MUSPSPAL	MUSPSPAL	284 bp	DNA	linear	ROD 20-MAY-1994
LOCUS					
DEFINITION	Mouse parotid secretory protein (PSP) gene, exon 1,				clone Y1E2.8.
ACCESSION	M26805				
VERSION	M26805.1	GI:200554			
KEYWORDS	parotid secretory protein; secretory protein.				
SEGMENT	1 of 3				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				





121 CCATCACTGTCACCAAGCATCCTGGGTGACCTTGGGTTGGAATATCTTGCTAACCTT 180

[illegible][illegible]



```

repeat_region      /rpt_family="L1"
                    28037..28164
repeat_region      /rpt_family="Alu"
                    30431..31273
repeat_region      /rpt_family="ERVK"
                    31535..31609
repeat_region      /rpt_family="B4"
                    31833..31922
repeat_region      /rpt_family="Alu"
                    31962..32291
repeat_region      /rpt_family="ERVK"
                    32338..33367
repeat_region      /rpt_family="ERVK"
                    37527..37856
repeat_region      /rpt_family="ERVK"
                    37975..38424
repeat_region      /rpt_family="RMR12"
                    38425..38774
repeat_region      /rpt_family="L1"
                    39296..39555
repeat_region      /rpt_family="RMR12"
                    40214..42773
repeat_region      /rpt_family="L1"
                    42977..43025
repeat_region      /rpt_family="L1"
                    43015..43516
repeat_region      /rpt_family="L1"
                    45272..45464
repeat_region      /rpt_family="B4"
                    51509..51627
repeat_region      /rpt_family="ERV1"
                    51836..52252
repeat_region      /rpt_family="L1"
                    52249..52290
repeat_region      /rpt_family="L1"
                    52806..52922
repeat_region      /note="Sequence derived from one plasmid subclone."
                    52864..52951
repeat_region      /rpt_family="L1"
                    52953..53119
repeat_region      /rpt_family="B2"
                    53160..53479
repeat_region      /rpt_family="L1"
                    53530..53636
repeat_region      /rpt_family="L1"
                    53637..54191
repeat_region      /rpt_family="ERVK"
                    54515..54872
repeat_region      /rpt_family="L1"
                    55568..55649
repeat_region      /rpt_family="B4"
                    56922..57044
repeat_region      /rpt_family="B4"
                    57783..57888
repeat_region      /rpt_family="B2"
                    58600..58763
repeat_region      /rpt_family="L1"
                    58764..58853
repeat_region      /rpt_family="ERVK"
                    64713..64893
repeat_region      /rpt_family="L1"
                    64951..65149
repeat_region      /rpt_family="L1"
                    65128..65337
repeat_region      /rpt_family="L1"
                    66707..66851
repeat_region      /rpt_family="ERV1"
                    66855..66923
repeat_region      /rpt_family="MALR"
                    67272..67468
repeat_region      /rpt_family="B2"
                    68460..68569

```

```

Query Match      35.0%; Score 175.6; DB 10; Length 165313;
Best Local Similarity 73.2%; Pred. No. 1.8e-43;
Matches 284; Conservative 0; Mismatches 89; Indels 15; Gaps 4;

QY      122 CATCACTGTCACCCAGACATCTGGGTGACCTTTGGTTGGAATATCTTGCTAACCTTA 181
         |||||
Db      46859 CATCAAAACCATTCATATGCTTGGGTGACCTTTGGTTGGAATATGTTGCT-ACCTTA 46917

QY      182 GATATTTCTTGGAGTCTCTTTGAACATCCAGAAATAGGGCTGATCTTCATCTCG 241
         |||||
Db      46918 GGTATTACTTGGACTCTTTTGAACATCTTGAACCAAGACTTGCTTCTCATCTCTG 46977

QY      242 GACCACATAT-----AAGTCAACCTTGAATCCAGAGATCGTCAGAGAAACAA 292
         |||||
Db      46978 GACCTCATATATATAAAGCCAAAGTACCTTGATTCAGATCCAGTGCTGATAAAAACAA 47037

QY      293 GGATCTCTCTGTGTGTCATCTCTTCAAGACATGATGACTCCACTTAACTGAGT 352
         |||||
Db      47038 GGA----TTTCATGTGTCATCATTTTCAACCTTCAATAGGGATCCCTGACTGTGA-C 47092

QY      353 TCCCATCTGAGAGTCACAGAGAGGCTTTGGGCAAGACAGAGGAAGCACTGTTGT 412
         |||||
Db      47093 ACACATCTGAGGGCTTTTGAAGAGCTTTAGAACAAAGACACAGGAAGCTTCACTTGT 47152

QY      413 GTTGTAAAGTTTGTACTCTTAACAATTTGAAGACATGATGACTGTGTCAACTAAC 472
         |||||
Db      47153 GTTGGGATGATATTATTAACATAAATTGAAGCATGATTACTGTGCAGACTAAT 47212

QY      473 AACAACTTACTCATGTGGGTTCTGTT 500
         |||||
Db      47213 GACAATCAAGTTTCATGTGTTCTGTT 47240

RESULT 13
AC147545/C      211446 bp      DNA      linear      ROD 08-SEP-2004
LOCUS      Mus musculus chromosome 14 clone RP23-355L20, complete sequence.
DEFINITION      AC147545
ACCESSION      AC147545.3 GI:51921462
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      The sequence of Mus musculus clone
JOURNAL      Unpublished
2 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      Direct Submision
JOURNAL      Submitted (10-DEC-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      Direct Submision
JOURNAL      Submitted (28-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      Direct Submision
JOURNAL      Submitted (08-SEP-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT
On Sep 8, 2004 this sequence version replaced gi:47777607.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: W_BA0355L20

```

FEATURES  
source  
Location/Qualifiers  
1..211446  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="14"  
/clone="RP23-355L20"

ORIGIN  
Query Match 35.0%; Score 175.6; DB 10; Length 211446;  
Best Local Similarity 73.2%; Pred. No. 1.7e-43;  
Matches 284; Conservative 0; Mismatches 89; Indels 15; Gaps 4;

QY 122 CATCACTGTCAACCAAGCATCTGGGACCTTTGGGTTGGAATATCTGGCTTAACCTTA 181  
197638 CATCAAAACCATTCATAGCTTGGGACCTTTGGGTTGGAATATGTTGGCT-ACCTTA 197580  
DB 182 GATATTTCTTTGAGATCTTTAGAACATCCAGAAATAGGGCTTGATCTCATCTCG 241  
197579 GGTATTTACTTGGACTCTTTTGAACATCTGTGACCAAGACTTGCCTTCTCATCTGT 197520  
QY 242 GACCAACAATAT-----AGTCACTTGAATCCAGAGATCGTCGAGAAACAA 292  
DB 197519 GACCTCAATATATAAGCAAGTCACTTGAATTCAGGTGCTGAGTAAAAACAA 197460  
QY 293 GATCTCTCTGTCGTCATCTTCTTCAAGCACTGAGTGTGACTCTCACTAACTGAGT 352  
DB 197459 GGA-----TTTCATGTGATATTTTCAAACTTCAAAATAGGGGATCTCTCATCTGTA-C 197405  
QY 353 TCCATCTGAGAGCCAGAGGCTTTGGGGCAAGAGAGAGGAGGACCTGTTGT 412  
DB 197404 ACAATCTGAGGGCTTCAAGAGCTTTAGAACAGAGACAGGAGGTTCACTTTGT 197345  
QY 413 GTTGTAAGTTTGACTTACCAAAATTTGAAGACATAGATGACATTTGTCTGACACTAC 472  
DB 197344 GTTGGGATGATATTATTAACATACAAATTTGAAGGATGATTAATGCTGTGACACTAAT 197285  
QY 473 AACCACTTGAAGCTATGGGTTCTGTT 500  
DB 197284 GACATCAAGGTTGATGTTGTTCTGTT 197257

RESULT 14  
AF153355  
LOCUS  
DEFINITION  
Rattus norvegicus Smgb gene, upstream sequence and exon 1.  
ACCESSION  
AF153355  
VERSION  
AF153355.1 GI:5031464  
KEYWORDS  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1196)  
REFERENCE  
AUTHORS  
Gupta,N., Asp,E., Levan,G. and Mirels,L.  
TITLE  
Structure and chromosomal localization of the rat salivary Psp and  
Smgb genes  
JOURNAL  
Gene 243 (1-2), 11-18 (2000)  
MEDLINE  
20156364  
PUBMED  
10675608  
2 (bases 1 to 1196)  
REFERENCE  
AUTHORS  
Mirels,L.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (21-MAY-1999) Molecular and Cell Biology, University of  
California, Berkeley, 401 Barker Hall #3204, Berkeley, CA  
94720-3204, USA  
FEATURES  
source  
Location/Qualifiers  
1..1196  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"

ORIGIN  
Query Match 33.0%; Score 165.2; DB 10; Length 1196;  
Best Local Similarity 76.9%; Pred. No. 3.9e-40;  
Matches 260; Conservative 0; Mismatches 58; Indels 20; Gaps 4;

QY 3 CCCACAGAACATCTTCCATGATTCATCTGATCTTCTTCTTGAACAATTCGAAC 62  
DB 864 CACACACAAATTCCTCCATTCATCTGATCTTCTTCTTCCAGACAAATTCGAAC 923  
QY 63 GACCCCTTGGCACTCTCAAGTTGACACATAGGGCTTG-----AAGTGGCA 114  
DB 924 GA-CACATCTGGACCTCAGCTCAGTCTTCACTCAGAAATCACTTGGCAGAACTTCT 982  
QY 115 CAAGTTCATCACTGTCAACCAAGCATCTGGGTGAACCTTTGGTTGAATATCTTGCT 174  
DB 983 CTCGATCCATCATGTGTCATCAAGCATCATGTGGTGAACCTTTGATTTGATATCTTGCT 1042  
QY 175 AACCTTAGATATTTCTTTGAGATCTTTAGAACATCCAGAAATAGGGCTTGAATCTC 234  
DB 1043 AACTTAGGATATTTCTTTGAGATCTTTAGAACATTCGGAATAGGGCTTGAATCTC 1102  
QY 235 ATCTTGGAGCCAAATATA-----GTCACTTGAATCCAGAGATCTGTCAGAGA 287  
DB 1103 ATTTGGAGACCAATATAAGGCCAGGTCACTTGAATCCAGAGATCTGTAAGAGA 1162  
QY 288 AACAGGATCTCTGTCATCTTCTTCAAGCA 325  
DB 1163 AACAGG-----CTTCTGTCTCTTCTTCCAAACA 1196

RESULT 15  
HSBA49G10  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP11-49G10 on chromosome 20. Contains  
a gene similar to bovine salivary protein BSP30, the LOC51297 gene  
for LUNX protein, 5' end of a gene encoding a protein similar to  
murine von Ebner minor salivary gland protein, a novel gene, a  
pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene  
similar to ribosomal protein L12, a putative novel transcript,  
ESTs, STSs, GSSs and a CpG island, complete sequence.  
ACCESSION  
AL121901  
VERSION  
AL121901.20 GI:8249854  
KEYWORDS  
HTG; BSP30; CpG island; LOC51297; LUNX; RPL12; STAT-inhibitor.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 161593)  
REFERENCE  
AUTHORS  
Tracey,A.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
request: clonerequest@sanger.ac.uk  
On Jun 5, 2000 this sequence version replaced gi:8017404.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
EM, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEB; Information

on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-49G10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-49G10 is at 1 in this sequence. The true left end of clone RP5-1187J4 is at 161494 in this sequence. The true right end of clone RP4-7330J3 is at 27823 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-49G10 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

```
FEATURES
source
    1..161593
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="20"
        /clone="RP11-49G10"
        /clone_1fb="RPCI-11.1"
        /complement(1..107)
        /note="match: GSS: Em:AQ607627"
        6..706
        /note="match: GSS: Em:AQ051854"
        137..548
        /note="match: GSS: Em:AQ550113"
        530..603
        /note="37 copies 2 mer ca 66% conserved"
        533..686
        /note="2 copies 77 mer 83% conserved"
        613..678
        /note="33 copies 2 mer ca 75% conserved"
        3030..3174
        /note="LIPB3 repeat: matches 5941..6084 of consensus"
        5140..5183
        /note="22 copies 2 mer tt 75% conserved"
        5184..5344
        /note="LIPB2 repeat: matches 5986..6146 of consensus"
        /complement(6104..6463)
        /note="match: GSS: Em:AQ339300"
        6120..6239
        /note="2 copies 60 mer 85% conserved"
        /complement(8476..9002)
        /note="match: GSS: Em:AQ537190"
        /complement(14766..15061)
        /note="match: STS: Em:HS922114S"
        14772..14779
        /note="Tandem repeat. Forced join. Approximately 500 bases missing according to restriction digest."
        16219..16387
        /note="match: GSS: Em:AL001671"
        16225..16387
        /note="match: GSS: Em:AL013146"
        18325..18826
        /note="match: GSS: Em:AQ458133"
        18328..18828
        /note="match: GSS: Em:AQ53957"
        18333..18770
        /note="match: GSS: Em:AQ765930"
        /complement(20975..21245)
        /note="match: GSS: Em:A2102442"
        20999..21238
```

```
misc_feature
    21031..21263
        /note="match: GSS: Em:A2055779 Em:A2084220"
        21144..21225
        /note="41 copies 2 mer ag 75% conserved"
        /complement(21172..21428)
        /note="match: GSS: Em:AQ395327"
        /complement(23473..23950)
        /note="match: GSS: Em:AQ727877"
        /complement(23637..23948)
        /note="match: GSS: Em:AQ691469"
        23950..24335
        /note="match: GSS: Em:AQ152593"
        23955..24456
        /note="match: GSS: Em:AQ508832"
        26979..27422
        /note="match: GSS: Em:AQ636646"
        31882..32061
        /note="3 copies 60 mer 73% conserved"
        31921..32032
        /note="56 copies 2 mer ct 63% conserved"
        33693..33762
        /note="MER47 repeat: matches 2251..2322 of consensus"
        33733..33785
        /note="MER47 repeat: matches 1..55 of consensus"
        35535..48787
        /note="DA49G10.1"
        join(35535..35585,36507..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47936)
        48600..48787
        /gene="DA49G10.1"
        /product="DA49G10.1 (similar to bovine salivary protein BSP30)"
        /note="match: CDNAS: Em:U79414 Em:U79413 Em:AX061621"
        match: ESTs: Em:A1654622"
        /evidence=not_experimental
        join(36522..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47936)
        /gene="DA49G10.1"
        /note="match: proteins: Tr:p79125 Sw:p07743"
        /codon_start=1
        /evidence=not_experimental
        /product="DA49G10.1 (similar to bovine salivary protein BSP30)"
        /protein_id="CAC03546.1"
        /db_xref="UniProt/TrEMBL:O9B000"
        /db_xref="UniProt/TrEMBL:O9B000"
        /translation="MLQMKVLLCGVITGSESLDNLGNDISVNDKLEPVLHEGL
        EIVDYLKIGLEKLVKVDGLVQKSSAWQLAKQAEAEKLNNTSKILPTTDFGL
        KINSLLIDVKAEPIDDGKGLNLSPPVAVNVAPPIIGIINLKASLDLPAVETP
        DQTHQPVAVLGECAEDPTSLISLIDKHSQINRFVNSVINTLSTVSSILQKEICP
        LIRIRHSDVNVIQVVDNPHKTLQDTLI"
        38155..38751
        /note="LIPB8 repeat: matches 4514..5125 of consensus"
        38750..39785
        /note="LIPB8 repeat: matches 5128..6163 of consensus"
        48769..48774
        /gene="DA49G10.1"
        48787
        /gene="DA49G10.1"
        52540..53450
        /note="CpG island"
        /evidence=not_experimental
        53831..54141
        /note="match: STS: Em:G31731"
        /complement(53831..54131)
        /note="match: GSS: Em:AQ897745"
        53842..53969
        /note="match: GSS: Em:AQ429347"
        /complement(53844..54141)
        /note="match: GSS: Em:AP101963"
        53858..54022
        /note="match: GSS: Em:AQ392891"
```

```
misc_feature      complement(53908..54141)
                   /note="match: GSS: Em:AQ415287"
misc_feature      53958..54139
                   /note="match: GSS: Em:AQ378081"
misc_feature      53970..54139
                   /note="match: GSS: Em:AQ527291"
misc_feature      53972..54142
                   /note="match: GSS: Em:AQ429347"
misc_feature      54000..54141
                   /note="match: GSS: Em:AQ392510"
misc_feature      complement(54027..54139)
                   /note="match: GSS: Em:AQ533658"
misc_feature      54030..54138
                   /note="match: GSS: Em:AG000935"
misc_feature      54041..54149
                   /note="match: GSS: Em:AQ392891"
gene              54767..55668
                   /gene="bA49G10.2"
CDS               /pseudo
                   54767..55668
                   /gene="bA49G10.2"
                   /note="bA49G10.2 (similar to STAR-induced STAR
inhibitor-2)
```

Query Match 10.5%; Score 52.4; DB 9; Length 161593;  
Best Local Similarity 60.1%; Pred. No. 5.2e-05;  
Matches 197; Conservative 0; Mismatches 91; Indels 40; Gaps 5;

```
QY 135 CAAGCATCTGGGTGACCTTGGGTGGAATATCTTGCTAACCTTA-----G 182
Db 35364 CAAGCATCTGGGTGACCTTGGGTGGAATATCTCTCTGCTCATCTCGAGAGGGG 35423
QY 183 ATATTTCTTGGAGTATCTTTAGACATCCAGGAAATAGGCTTGATCT----- 233
Db 35424 TTATTTCTTGGAGTATCTTTAGATGATGCCAGAACCAAGGCTTGCTGCGGAGC 35483
QY 234 ----CATCTGGGACCAATATATAGTACCC-----TGAATCCAGAGATCG 279
Db 35484 CTTCATCTCCATCTCCACATATATAGCCGCCACATTCAGCAGAAATCCAGGACTG 35543
QY 280 TGCAGAGAAACAAGATCTCTCGTGTGATCTCTTCAAAAGCAGTAGTAGTCACTC 339
Db 35544 TGCAGTGGGGCAAGA---TTTCATAGCATCTCTCTTAACCGCTGAGTGGCGTCT 35599
QY 340 CACTAACTGAGTTCATCTGAGATCCACAGAGGCTTT-GGAGCAAGAGCAGAGGG 398
Db 35600 CCCCCGATGCGACACACGTCGTCTTTGGAAGGCTTTGGGGCAAGAGCAGAGGG 35659
QY 399 AAGGACTGTTTGTGTGTAAGTTT 426
Db 35660 GTGGGATTGCTTGGGTAGCTAATATTT 35687
```

Search completed: February 10, 2005, 14:09:34  
Job time : 2418.78 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33 ; Search time 2171.78 Seconds  
(without alignments)  
8780.927 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_11500\_12000

Perfect score: 501  
Sequence: 1 aaccacaacagacatcttc.....gactcatcgggttcgttc 501

Scoring table: IDENTITY NUC  
Gapop 10\_0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	25.3	747	4	BG865470 602783596
2	127	25.3	865	4	BG871618 602790109
3	127	25.3	866	4	BG869398 602788759
4	126	25.1	767	4	BG867661 602787451
5	125	25.0	680	4	BM171830 imagegc 5
6	125	25.0	744	4	BG865845 602788261
7	125	25.0	938	4	BG867523 602788460
8	124	24.8	832	4	BG869852 602789378
9	122	24.4	1007	4	BG867502 602788430
10	120	24.0	754	4	BG873122 602794520
11	119	23.8	249	1	AA562244 602783777
12	117.6	23.5	871	6	CB575686 AGENCOURT
13	115	23.0	563	8	CB575686 AGENCOURT
14	115	23.0	671	2	BF301350 602029803
15	115	23.0	827	2	BF301350 602029803
16	114.4	22.8	442	1	AA798080 vx67a05.r
17	113	22.6	686	4	BG867018 602786008
18	113	22.6	920	4	BG865620 602783777
19	108	21.6	918	2	BF30134 602032028
20	107.6	21.5	880	2	BF300134 602032028
21	107.4	21.4	955	2	BF540271 602053225
22	105.4	21.0	771	2	BF537617 602053392
23	105	21.0	848	4	BG869732 602787559
24	104	20.8	903	2	BF302106 602033055

25	103	20.6	726	2	BF302348
26	99.4	19.8	781	2	BF538311
27	99	19.8	539	1	AA930330
28	99	19.8	844	2	BF300497
29	89	17.8	850	2	BF302185
30	82.8	16.5	873	2	BF535056
31	80.4	16.0	365	9	CR145142
32	74	14.8	797	1	AA726504
33	57	11.4	813	2	BF306339
34	57	11.4	996	2	BF301924
35	55	11.0	843	4	BG870730
36	49	9.8	784	4	BG871734
37	49	9.8	992	4	BG870291
38	47.6	9.5	428	6	AI930435
39	47	9.4	395	1	AI930435
40	47	9.4	357	6	BY776189
41	47	9.4	972	3	AK009032
42	47	9.4	972	3	AK010181
43	46	9.2	461	1	AI574478
44	46	9.2	625	1	AI151916
45	44	8.8	1046	4	BG869150

## ALIGNMENTS

RESULT 1  
LOCUS BG865470 747 bp mRNA EST 29-MAY-2001  
DEFINITION 602783596F1 NCI\_CGAP\_SG2 Mus musculus CDNA clone IMAGE:491062 5',  
mRNA sequence.

ACCESSION BG865470 GI:14216010  
VERSION BG865470.1  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus (house mouse)  
EST.  
Mus musculus

REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10809 row: 0 column: 23  
High quality sequence start: 2  
High quality sequence stop: 746.  
Location/Qualifiers

## FEATURES

1..747  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:491062"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: salivary gland, Vector: pCMV-SPORT6, Site: 1:  
Nci1, Site 2: Sal1; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 25.3%; Score 127; DB 4; Length 747;  
Best Local Similarity 100.0%; Pred. No. 3.3e-07;  
Matches 127, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

email: cgadams-rw@mail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 plate: LLNM10839 row: b column: 04  
 high quality sequence stop: 809.

Location/Qualifiers  
1. .865

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:491251"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NTL_GCAP_S62"
/notes="Organ: salivary gland. Vector: pCMV-Sport6, Site 1: NotI, Site 2: SalI, Cloned unidirectionally. Primer: Oligo NotI. Average insert size 1.3 Kb. Constructed by life technologies. Note: this is a NTL_GCAP Library."

```

25.3%; Score 127; DB 4; Length 865;

Query Match	25.3%	Score 127	DB 4	length 865
Best Local Similarity	100.0%	Prod. No.	3 4e-27	
Matches 127	0	Mismatches	0	Gaps 0

QY	290	CAAGATCTCTCTCGTGTGATCTCTTCTTCAAGCAGTGAAGTACTCCATCTAAAC	349
Db	1	CAAGATCTCTCTCGTGTGATCTCTTCTTCAAGCAGTGAAGTACTCCATCTAAAC	60
QY	350	AGTTCCTCATCTGAGAGTCCAAGAGAGGCTTTTGGGGCAAGAAACAGAGGGAAGGCACTGTT	409
Db	61	AGTTCCTCATCTGAGAGTCCAAGAGAGGCTTTTGGGGCAAGAAACAGAGGGAAGGCACTGTT	120

QY	410	TGTGTTG	416
Db	121	TGTGTTG	127

### RESULT 3

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4920182"
/lab_host="DHIO8 (T1 phage-resistant)"
/clone_lib="NCI CGAP_S62"
/notes="Organ: salivary gland; Vector: pCMV-SPORE6; Site 1: Not1; Site 2: Sal1; Cloned unidirectionally. Primer: Oligo dt1. Average insert size 1.3 Kb. Constructed by Life technologies. Note: this is a NCI CGAP Library."

```

25.38; Score 127; DB 4; Length 866;

Query Match	25.3%	Score 127;	DB 4;	Length 866;
Best Local Similarity	100.0%	Pred. No. 3.5e-27;		
Matches 127;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0.

**QY** 290 CAAGATCTCTCGTGTGCATCCTTTCTTCAAGACGAGTAGTCATCCACTAACTG 349  
|||||  
**Dd** 13 CAAGATCTCTCTCGTGTGCATCCTTTCTTCAAGACGAGTAGTCATCCACTAACTG 72  
|||||

Oy	350	AGTTCATCTGAGAGTCCACAGAGGCTTTGGGGCAGAAACAGAGGGAGGCACTGTT	409
Dh	73	AGTTCATCTGAGAGTCCACAGAGGCTTTGGGGCAGAAACAGAGGGAGGCACTGTT	132

QY	410	TGTGTTG	416
Db	133	TGTGTTG	139

767 bp	mRNA	linear	EST 29-MAY-2003

LOCUS	767 bp	mRNA	linear	EST 29-MAY-2001
DEFINITION	BC867661			
	602787451	NCI_CGAP_SG2	Mus musculus	CDNA clone IMAGE:4913292 5'
	mRNA sequence.			

ACCESSION	BG867661	GI:14218201
VERSION	BG867661.1	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 767)  
NIH-MGC <http://mgc.ncl.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)

Page 2

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM10818 row: f column: 13  
High quality sequence stop: 759.  
Location/Qualifiers

## FEATURES

## source

1. .767  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4913292"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site: 1;  
NCI; Site\_2: Sal; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

## Query Match

25.1%; Score 126; DB 4; Length 767;  
Best Local Similarity 100.0%; Pred. No. 6.7e-27; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 AAGATCTCTCTGTCGATCCCTTTCTTCAAGCAGTAGTACTCCATAACTGA 350

Db 1 AAGATCTCTCTGTCGATCCCTTTCTTCAAGCAGTAGTACTCCATAACTGA 60

QY 351 GTTCCATCTGAGAGTCCACAGAGGCTTTGGGCAAGACAGAGGAGCACTGTTT 410

Db 61 GTTCCATCTGAGAGTCCACAGAGGCTTTGGGCAAGACAGAGGAGCACTGTTT 120

QY 411 GTGTTG 416

Db 121 GTGTTG 126

## RESULT 5

## BM171830

LOCUS 680 bp mRNA linear EST 04-DEC-2001  
DEFINITION imagec\_5\_2001/ana193bdr1.y1 NCI\_CGAP\_SG2 Mus musculus cDNA clone  
IMAGE:4922760 5', mRNA sequence.

ACCESSION BM171830  
VERSION BM171830.1 GI:17311393

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 680)

REFERENCE Kale, P.I., Harsch, T.J., Folta, P.A., Nelson, D.O., Sanders, C.G. and

Prange, C.K. The I.M.A.G.E. Consortium quality control effort: clone

resequencing for verification

Unpublished (2001)

Contact: Prange CK

The I.M.A.G.E. Consortium

Lawrence Livermore National Laboratory

Livermore, CA, USA

Email: help@image.jnl.gov

This read has been produced as part of the I.M.A.G.E. Consortium  
quality control effort. High quality sequence is defined as having  
100 or more base pairs with a phred quality value of 20 or greater,  
where a sliding window of 4 base pairs with a phred quality value  
of 15 or greater marks the beginning and end of the sequence. For  
information on obtaining this clone, please contact

## FEATURES

## source

info@image.jnl.gov.  
Plate: LLM10843 row: a column: 1  
Seq primer: m3rp1  
High quality sequence stop: 680.  
Location/Qualifiers

1. .680  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4922760"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site: 1;  
NCI; Site\_2: Sal; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

## Query Match

25.0%; Score 125; DB 4; Length 680;  
Best Local Similarity 100.0%; Pred. No. 1.3e-26; Indels 0; Gaps 0;

Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AAGATCTCTCTGTCGATCCCTTTCTTCAAGCAGTAGTACTCCATAACTGAG 351

Db 1 AAGATCTCTCTGTCGATCCCTTTCTTCAAGCAGTAGTACTCCATAACTGAG 60

QY 352 TTCCATCTGAGAGTCCACAGAGGCTTTGGGCAAGACAGAGGAGCACTGTTT 411

Db 61 TTCCATCTGAGAGTCCACAGAGGCTTTGGGCAAGACAGAGGAGCACTGTTT 120

QY 412 TGTTG 416

Db 121 TGTTG 125

## RESULT 6

## BG865845

LOCUS 744 bp mRNA linear EST 29-MAY-2001  
DEFINITION 602788261F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4914158 5',  
mRNA sequence.

ACCESSION BG865845  
VERSION BG865845.1 GI:14216385

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 744)  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.jnl.gov

Plate: LLM10820 row: j column: 15

High quality sequence stop: 708.

Location/Qualifiers

## FEATURES

## source

1. .744  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4914158"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site: 1;

## ORIGIN

Note: Site 2: Salt; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

Query Match 25.0%; Score 125; DB 4; Length 744;  
Best Local Similarity 100.0%; Pred. No. 1.3e-26;  
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AGAATCTCTCTGTCATCTCTTCTTCAAGCAGTAGTAGTCACTCCACTAACTGAG 351  
|||||  
1 AGAATCTCTCTGTCATCTCTTCTTCAAGCAGTAGTAGTCACTCCACTAACTGAG 60  
DB 352 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGGCACTGTTTG 411  
|||||  
61 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGGCACTGTTTG 120  
QY 412 TGTGTG 416  
|||||  
DB 121 TGTGTG 125

RESULT 7  
BG867523 938 bp mRNA linear EST 29-MAY-2001  
LOCUS 602788460F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4919847 5',  
DEFINITION mRNA sequence.  
ACCESSION BG867523  
VERSION BG867523.1 GI:14218063  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 938)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Note="Organ: salivary gland; Vector: PCMV-SPORE6; Site 1:  
Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

FEATURES  
source  
1..938  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4919847"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: salivary gland; Vector: PCMV-SPORE6; Site 1:  
Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
High quality sequence stop: 812.  
Location/Qualifiers

ORIGIN  
Query Match 25.0%; Score 125; DB 4; Length 938;  
Best Local Similarity 100.0%; Pred. No. 1.4e-26;  
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AGAATCTCTCTGTCATCTCTTCTTCAAGCAGTAGTAGTCACTCCACTAACTGAG 351  
|||||  
1 AGAATCTCTCTGTCATCTCTTCTTCAAGCAGTAGTAGTCACTCCACTAACTGAG 60  
DB 352 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGGCACTGTTTG 411

DB 61 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGGCACTGTTTG 120  
|||||  
QY 412 TGTGTG 416  
|||||  
DB 121 TGTGTG 125

RESULT 8  
BG869852 832 bp mRNA linear EST 29-MAY-2001  
LOCUS 602789378F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4920739 5',  
DEFINITION mRNA sequence.  
ACCESSION BG869852  
VERSION BG869852.1 GI:14220392  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 832)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Note="Organ: salivary gland; Vector: PCMV-SPORE6; Site 1:  
Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
High quality sequence stop: 807.  
Location/Qualifiers

FEATURES  
source  
1..832  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4920739"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone="NCI\_CGAP\_SG2"  
/note="Organ: salivary gland; Vector: PCMV-SPORE6; Site 1:  
Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
High quality sequence stop: 807.  
Location/Qualifiers

ORIGIN  
Query Match 24.8%; Score 124; DB 4; Length 832;  
Best Local Similarity 100.0%; Pred. No. 2.8e-26;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 GATCTCTCTGTCATCTCTTCTTCAAGCAGTAGTAGTCACTCCACTAACTGAGT 352  
|||||  
11 GATCTCTCTGTCATCTCTTCTTCAAGCAGTAGTAGTCACTCCACTAACTGAGT 70  
DB 353 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGGCACTGTTTG 412  
|||||  
71 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGGCACTGTTTG 130  
QY 413 GTTGTG 416  
|||||  
DB 131 GTTGTG 134

RESULT 9  
BG867502 1007 bp mRNA linear EST 29-MAY-2001  
LOCUS 602788460F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4919955 5',  
DEFINITION mRNA sequence.  
ACCESSION BG867502

VERSION BG67502.1 GI:14218042  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1007)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 Tissue Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLAM10835 row: k column: 08  
 High quality sequence stop: 801.  
 Location/Qualifiers  
 1..1007  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4919935"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_1lb="NCI CGAP SG2"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."

# FEATURES

Query Match 24.4%; Score 122; DB 4; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-25;  
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## ORIGIN

295 ATCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCACTAACTGATTC 354  
 1 ATCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCACTAACTGATTC 60  
 355 CCATCTGAGATCCACAGAGGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 414  
 61 CCATCTGAGATCCACAGAGGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 120  
 415 TG 416  
 121 TG 122

RESULT 10  
 BG873122 754 bp mRNA linear EST 29-MAY-2001  
 LOCUS 602794520P1 NCI CGAP\_SG2 Mus musculus cDNA clone IMAGE:4925621 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG873122  
 VERSION BG873122.1 GI:14223662  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 754)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLAM10850 row: h column: 06  
 High quality sequence stop: 754.  
 Location/Qualifiers  
 1..754  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4925621"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_1lb="NCI CGAP SG2"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."

## ORIGIN

Query Match 24.0%; Score 120; DB 4; Length 754;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-25;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CTCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCACTAACTGATTC 356  
 DB 13 CTCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCACTAACTGATTC 72  
 QY 357 ATCTGAGATCCACAGAGGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 416  
 DB 73 ATCTGAGATCCACAGAGGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 122

RESULT 11  
 AA562244 249 bp mRNA linear EST 18-AUG-1997  
 LOCUS V134B05.r1 Stragene mouse skin (#937311) Mus musculus cDNA clone  
 DEFINITION IMAGE:974097 5' similar to gp:X01697 Mouse mRNA fragment for parotid secretory protein (MOUSE); mRNA sequence.  
 AA562244  
 ACCESSION AA562244.1 GI:2333709  
 VERSION  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 249)  
 Marris, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheilenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 The Washu-HMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Maira W/Mouse EST Project  
 Washu-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:554825  
 Seq primer: -26ml3 rev1 ET from Amersham  
 High quality sequence stop: 241.  
 Location/Qualifiers  
 1..249  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"

## FEATURES

source

/db\_xref="taxon:10090"  
/clone="IMAGE:974097"  
/sex="females"  
/tissue\_type="whole skin"  
/dev stage="11 weeks old"  
/lab host="SOLR (kanamycin resistant)"  
/clone\_11b="Stratagene mouse skin (#937313)"  
/note="Organ: skin; Vector: pBluescript SK<sup>+</sup>; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: 011go dr. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'."

## ORIGIN

Query Match 23.8%; Score 119; DB 1; Length 249;  
Best Local Similarity 100.0%; Pred. No. 6.7e-25;  
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 TCTCTGCTGTCATCTTTTCAAGCAGTGTAGTACTCTCACTAACTGAGTTCCCA 357  
| | | | |  
DB 1 TCTCTGCTGTCATCTTTTCAAGCAGTGTAGTACTCTCACTAACTGAGTTCCCA 60  
| | | | |  
QY 358 TCTGAGAGTCCACAGAGGCTTTGGGGCAAGAGCAGAGGAGGACCTGTTGTGTTG 416  
| | | | |  
DB 61 TCTGAGAGTCCACAGAGGCTTTGGGGCAAGAGCAGAGGAGGACCTGTTGTGTTG 119  
| | | | |

## RESULT 12

CB575686 871 bp mRNA linear EST 02-APR-2003  
LOCUS CB575686  
DEFINITION IMAGE:30292464 5', mRNA sequence.

ACCESSION CB575686  
VERSION CB575686.1 GI:29495216  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
plate: NDAM36 row: 9 column: 01  
High quality sequence start: 30  
High quality sequence stop: 528.

## FEATURES

## source

1. 871  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30292464"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5' 4 and 1 limb and jaw equivalents from respective days). Cloned unidirectionally, oligo-dr primed (5'-GACTAGCTTAGATCGAGCGGCCGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by

## ORIGIN

Query Match 23.5%; Score 117.6; DB 6; Length 871;  
Best Local Similarity 96.8%; Pred. No. 2.6e-24;  
Matches 120; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 223 GAGTCTCTCTGTCATCTTTTCAAGCAGTGTAGTACTCTCACTAACTGAGT 352  
| | | | |  
DB 6 GAGTCTCTCTGTCATCTTTTCAAGCAGTGTAGTACTCTCACTAACTGAGT 65  
| | | | |  
QY 353 TCCCATCTGAGATCCACAGAGGCTTTGGGGCAAGAGCAGAGGAGGACCTGTTGT 412  
| | | | |  
DB 66 TCCCATCTGAGATCCACAGAGGCTTTGGGGCAAGAGCAGAGGAGGACCTGTTGT 125  
| | | | |

QY 413 GTTG 416  
| | | | |  
DB 126 GTTG 129

## RESULT 13

AO974026 563 bp DNA linear GSS 28-JAN-2000  
LOCUS AO974026  
DEFINITION RPCI-23-331F18, TV RPCI-23 Mus musculus genomic clone  
AO974026 genomic survey sequence.

ACCESSION AO974026.1 GI:6804479  
VERSION AO974026  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatman, S., Akincic, B., Levins, M., McGinn, S., Tesgaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other GSSs: RPCI-23-331F18.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhao@tigr.org](mailto:szhao@tigr.org)

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong ([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page: [http://www.tigr.org/cdb/bac/ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac/ends/mouse/bac_end_intro.html)  
Plate: 331 row: F column: 18  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

## source

1. 563  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-331F18"  
/sex="Female"  
/lab\_host="DH10B"  
/clone\_11b="RPCI-23"  
/note="Organ: Kidney/Brain; Vector: pAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pAC3.6 vector at the DH10B sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."



Fri Feb 11 08:00:44 2005

us-09-926-375b-7\_copy\_11500\_12000.rst

Page 8

Job time : 2175.78 secs

---





CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
 CC on 15-SEP-2003 to standardise OS field)

XX Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;

Query Match 100.0%; Score 501; DB 3; Length 17732;  
 Best Local Similarity 100.0%; Pred. No. 4e-151;  
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCACAGAACATCTTCCATGATCTGATCTGATCTTCTTGTAGACAAATCTGAA 60  
 DB AACCCACAGAACATCTTCCATGATCTGATCTGATCTTCTTGTAGACAAATCTGAA 11559  
 QY CTGACCCCTTGTGGAGCTCTCAAGTTTGACAACATAGGCTTTGAACCTTGACAAAGT 120  
 DB CTGACCCCTTGTGGAGCTCTCAAGTTTGACAACATAGGCTTTGAACCTTGACAAAGT 11619  
 QY 121 CCATCACTGTACCCCAAGATCTGAGTCTTGGGTTGGAATCTTGGCTAACCTT 180  
 DB 11620 CCATCACTGTACCCCAAGATCTGAGTCTTGGGTTGGAATCTTGGCTAACCTT 11679  
 QY 181 AGATATTTCTTGGAGTATCTTGAACATCCAGGAATAGGGCTTGATCTCATCTG 240  
 DB 11680 AGATATTTCTTGGAGTATCTTGAACATCCAGGAATAGGGCTTGATCTCATCTG 11739  
 QY 241 GGACCAATATTAAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 300  
 DB 11740 GGACCAATATTAAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 11799  
 QY 301 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 360  
 DB 11800 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 11859  
 QY 361 GAGAGTCCACAGAGGCTTGGGGCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
 DB 11860 GAGAGTCCACAGAGGCTTGGGGCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11919  
 QY 421 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGACATTAACAACACT 480  
 DB 11920 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGACATTAACAACACT 11979  
 QY 481 AGACTCATGTGGGTTCTGTTT 501  
 DB 11980 AGACTCATGTGGGTTCTGTTT 12000

## RESULT 2

AAC68294  
 ID AAC68294 standard; DNA; 20623 BP.

XX AAC68294;  
 AC 15-SEP-2003 (revised)  
 DT 20-FEB-2001 (first entry)  
 XX  
 DE Lama2/APPA plasmid coding sequence.  
 XX  
 KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KM environmental pollution; pig; ds.  
 OS Mus musculus.  
 OS Escherichia coli.  
 OS Chimeric.  
 XX  
 XX WO20064247-A1.  
 XX 02-NOV-2000.  
 XX 20-APR-2000; 2000WO-CA000430.  
 XX 23-APR-1999; 99US-0130508P.

XX (UYGU-) UNIV GUELPH.  
 PA Forsberg CW, Golovan S, Phillips JP;  
 PI MPI; 2000-687245/67.  
 DR P-PSDB; AAB36257.

XX Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.

PS Claim 56; Fig 5; 152pp; English.

XX The present invention provides transgenic animals which produce desired  
 CC proteins, in this case pigs which expresses phytase in the salivary  
 CC gland. Low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
 CC on 15-SEP-2003 to standardise OS field)

XX Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match 100.0%; Score 501; DB 3; Length 20623;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-151;  
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCACAGAACATCTTCCATGATCTGATCTGATCTTCTTGTAGACAAATCTGAA 60  
 DB 11500 AACCCACAGAACATCTTCCATGATCTGATCTGATCTTCTTGTAGACAAATCTGAA 11559  
 QY 61 CTGACCCCTTGTGGAGCTCTCAAGTTTGACAACATAGGCTTTGAACCTTGACAAAGT 120  
 DB 11560 CTGACCCCTTGTGGAGCTCTCAAGTTTGACAACATAGGCTTTGAACCTTGACAAAGT 11619  
 QY 121 CCATCACTGTACCCCAAGATCTGAGTCTTGGGTTGGAATCTTGGCTAACCTT 180  
 DB 11620 CCATCACTGTACCCCAAGATCTGAGTCTTGGGTTGGAATCTTGGCTAACCTT 11679  
 QY 181 AGATATTTCTTGGAGTATCTTGAACATCCAGGAATAGGGCTTGATCTCATCTG 240  
 DB 11680 AGATATTTCTTGGAGTATCTTGAACATCCAGGAATAGGGCTTGATCTCATCTG 11739  
 QY 241 GGACCAATATTAAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 300  
 DB 11740 GGACCAATATTAAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 11799  
 QY 301 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 360  
 DB 11800 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 11859  
 QY 361 GAGAGTCCACAGAGGCTTGGGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
 DB 11860 GAGAGTCCACAGAGGCTTGGGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11919  
 QY 421 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGACATTAACAACACT 480  
 DB 11920 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGACATTAACAACACT 11979  
 QY 481 AGACTCATGTGGGTTCTGTTT 501  
 DB 11980 AGACTCATGTGGGTTCTGTTT 12000

## RESULT 3

AAK89798  
 ID AAK89798 standard; DNA; 7524 BP.

XX AAK89798;  
 AC 05-NOV-2001 (first entry)  
 XX  
 DT

XX		Human digestive system antigenic genomic sequence SEQ ID NO: 3374.
XX	DE	Human, digestive system antigen; gene therapy; cancer; appendicitis;
KW	KM	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX	KX	digestive system disorder; Meckel's diverticulum; ds.
OS	Homo sapiens.	
PN	WO20015314-A2.	
PD	02-AUG-2001.	
PF	17-JAN-2001; 2001WO-US001324.	
PR	31-JAN-2000; 2000US-0179665P.	
PR	04-FEB-2000; 2000US-0180628P.	
PR	24-FEB-2000; 2000US-0184664P.	
PR	02-MAR-2000; 2000US-0186350P.	
PR	16-MAR-2000; 2000US-0189874P.	
PR	17-MAR-2000; 2000US-0190076P.	
PR	18-APR-2000; 2000US-0198123P.	
PR	19-MAY-2000; 2000US-0205515P.	
PR	07-JUN-2000; 2000US-0209467P.	
PR	28-JUN-2000; 2000US-0214886P.	
PR	30-JUN-2000; 2000US-0215135P.	
PR	07-JUL-2000; 2000US-0216647P.	
PR	07-JUL-2000; 2000US-0216880P.	
PR	11-JUL-2000; 2000US-0217487P.	
PR	11-JUL-2000; 2000US-0217486P.	
PR	14-JUL-2000; 2000US-0218230P.	
PR	26-JUL-2000; 2000US-0220963P.	
PR	26-JUL-2000; 2000US-0220964P.	
PR	14-AUG-2000; 2000US-0224518P.	
PR	14-AUG-2000; 2000US-0224519P.	
PR	14-AUG-2000; 2000US-0225213P.	
PR	14-AUG-2000; 2000US-0225214P.	
PR	14-AUG-2000; 2000US-0225266P.	
PR	14-AUG-2000; 2000US-0225267P.	
PR	14-AUG-2000; 2000US-0225268P.	
PR	14-AUG-2000; 2000US-0225270P.	
PR	14-AUG-2000; 2000US-0225447P.	
PR	14-AUG-2000; 2000US-0225757P.	
PR	14-AUG-2000; 2000US-0225758P.	
PR	14-AUG-2000; 2000US-0225759P.	
PR	18-AUG-2000; 2000US-0226279P.	
PR	22-AUG-2000; 2000US-0226681P.	
PR	22-AUG-2000; 2000US-0226686P.	
PR	22-AUG-2000; 2000US-0227182P.	
PR	23-AUG-2000; 2000US-0227009P.	
PR	30-AUG-2000; 2000US-0228924P.	
PR	01-SEP-2000; 2000US-0229287P.	
PR	01-SEP-2000; 2000US-0229343P.	
PR	01-SEP-2000; 2000US-0229345P.	
PR	05-SEP-2000; 2000US-0229359P.	
PR	05-SEP-2000; 2000US-0229513P.	
PR	06-SEP-2000; 2000US-0230438P.	
PR	06-SEP-2000; 2000US-0231242P.	
PR	08-SEP-2000; 2000US-0231243P.	
PR	08-SEP-2000; 2000US-0231244P.	
PR	08-SEP-2000; 2000US-0231413P.	
PR	08-SEP-2000; 2000US-0231414P.	
PR	08-SEP-2000; 2000US-0232080P.	
PR	12-SEP-2000; 2000US-0231968P.	
PR	14-SEP-2000; 2000US-0232397P.	
PR	14-SEP-2000; 2000US-0232398P.	
PR	14-SEP-2000; 2000US-0232399P.	
PR	14-SEP-2000; 2000US-0232400P.	
PR	14-SEP-2000; 2000US-0232401P.	
PR	14-SEP-2000; 2000US-0233063P.	
PR	14-SEP-2000; 2000US-0233064P.	
PR	14-SEP-2000; 2000US-0233065P.	
PR	21-SEP-2000; 2000US-0234274P.	
PR	25-SEP-2000; 2000US-0234997P.	
PR	25-SEP-2000; 2000US-0234998P.	
PR	26-SEP-2000; 2000US-0235484P.	
PR	27-SEP-2000; 2000US-0235834P.	
PR	27-SEP-2000; 2000US-0235836P.	
PR	29-SEP-2000; 2000US-0236327P.	
PR	29-SEP-2000; 2000US-0236367P.	
PR	29-SEP-2000; 2000US-0236368P.	
PR	29-SEP-2000; 2000US-0236369P.	
PR	29-SEP-2000; 2000US-0236370P.	
PR	02-OCT-2000; 2000US-0236802P.	
PR	02-OCT-2000; 2000US-0237037P.	
PR	02-OCT-2000; 2000US-0237038P.	
PR	02-OCT-2000; 2000US-0237039P.	
PR	02-OCT-2000; 2000US-0237040P.	
PR	13-OCT-2000; 2000US-0239935P.	
PR	13-OCT-2000; 2000US-0239937P.	
PR	20-OCT-2000; 2000US-0240960P.	
PR	20-OCT-2000; 2000US-0241221P.	
PR	20-OCT-2000; 2000US-0241785P.	
PR	20-OCT-2000; 2000US-0241786P.	
PR	20-OCT-2000; 2000US-0241787P.	
PR	20-OCT-2000; 2000US-0241808P.	
PR	20-OCT-2000; 2000US-0241809P.	
PR	20-OCT-2000; 2000US-0241826P.	
PR	01-NOV-2000; 2000US-0246474P.	
PR	08-NOV-2000; 2000US-0246475P.	
PR	08-NOV-2000; 2000US-0246476P.	
PR	08-NOV-2000; 2000US-0246477P.	</

XX		Mang L, Zamudio C, Malone C, Haeselbeck R, Ohlsen KL, Zyskind JW;
PI	Wall D, Trawick JB, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;	
DR	WPI: 2003-029926/02.	
DR	P-PSDB; ABU24464.	
PT	New antisense nucleic acids, useful for identifying proteins or screening	
PT	for homologous nucleic acids required for cellular proliferation to	
PT	isolate candidate molecules for rational drug discovery programs.	
XX		
PS	Claim 14; SEQ ID NO 16204; 1766bp; English.	
XX		
CC	The invention relates to an isolated nucleic acid comprising any one of	
CC	the 613 antisense sequences given in the specification where expression	
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:	
CC	(1) a vector comprising a promoter operably linked to the nucleic acid	
CC	encoding a polypeptide whose expression is inhibited by the antisense	
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated	
CC	polypeptide or its fragment whose expression is inhibited by the	
CC	antisense nucleic acid; (4) an antibody capable of specifically binding	
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
CC	proliferation or the activity of a gene in an operon required for	
CC	proliferation; (7) identifying a compound that influences the activity of	
CC	the gene product or that has an activity against a biological pathway	
CC	required for proliferation, or that inhibits cellular proliferation; (8)	
CC	identifying a gene required for cellular proliferation or the biological	
CC	pathway in which a proliferation-required gene or its gene product lies	
CC	or a gene on which the test compound that inhibits proliferation of an	
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a	
CC	compound's activity; (11) a culture comprising strains in which the gene	
CC	product is overexpressed or underexpressed; (12) determining the extent	
CC	to which each of the strains is present in a culture or collection of	
CC	strains; or (13) identifying the target of a compound that inhibits the	
CC	proliferation of an organism. The antisense nucleic acids are useful for	
CC	identifying proteins or screening for homologous nucleic acids required	
CC	for cellular proliferation to isolate candidate molecules for rational	
CC	drug discovery programs, or for screening homologous nucleic acids	
CC	required for proliferation in cells other than S. aureus, S. typhimurium,	
CC	K. pneumoniae or P. aeruginosa. The present sequence data for this patent did	
CC	not form part of the printed specification, but was obtained in	
CC	electronic format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 3555 BP; 1445 A; 358 G; 685 G; 1067 T; 0 U; 0 Other;	
	Query Match 7.4%; Score 37; DB 8; Length 3555;	
	Best Local Similarity 56.0%; Pred. No. 0.82;	
	Matches 70; Conservative 0; Mismatches 55; Indels 0; Gaps 0	
OY	181 AGATTATTTCTTGAGATCATCTTAAAGAATCAACGAGAATAAGGGCTGATTCATCCTG 240	
DB	1947 AGATTATTAATAGCGAATAAAGTTTATATAGACGAGGCCCTTAGTGAATAACCAAGTA 2006	
OY	241 GGACCACAATATTAAGTCAACCTAGAAATCCAGAGATGTGTGCAGAAAACAGATCTCT 300	
DB	2007 TATAGCTATATAAAATAAACCTTTAAATATATGAATATCAAAACAGAAAAACTAGAACCTAT 2066	
OY	301 CTCGT 305	
DB	2067 ATTGT 2071	
RESULT 5		
ACN43856/c		
ID	ACN43856 standard; DNA; 49088 BP.	
AC	ACN43856;	
XX		
DT	18-NOV-2004 (first entry)	
XX		
DE	Mouse genomic sequence MCG10112.	

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
XX Mus musculus.  
XX WO2003073826-A2.  
XX 12-SEP-2003.  
XX 28-FEB-2003; 2003WO-US006235.  
XX 01-MAR-2002; 2002US-00087192.  
XX (SAGR-) SAGRES DISCOVERY.  
XX Morris DW;  
XX WPI; 2003-328604/31.  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX  
XX Claim 1; SEQ ID NO 13; Opp; English.  
XX  
XX The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a block; (x)  
XX for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published  
XX  
XX Sequence 49088 BP; 13195 A; 9519 C; 10290 G; 14938 T; 0 U; 1146 Other;  
XX  
XX Query Match 7.3%; Score 36.8; DB 11; Length 49088;  
XX Best Local Similarity 59.6%; Pred. No. 3.2;  
XX Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
XX  
XX QY 255 GTCACCTAGATCCAGAGATGTCGAGAAACAAGATCTCTCGTGATCTCT 314  
XX DB 45517 GTCTCAAAACAACTGCTGGAGTGCACCACTGAAATTAAGTTTAAACCCCTGGAACT 45458  
XX QY 315 TCTTCAAGCAGTAGTAGTCACTCACTAACTGAGTTCCCAT 358  
XX DB 45457 ACATTAAGCAAGAAAGAAATGACCCCACTAAGTTAGTTTCAT 45414  
XX  
XX RESULT 6  
XX AAC68112  
XX ID AAC68112 standard; cDNA; 1470 BP.  
XX  
XX AAC68112;  
XX  
XX 20-FEB-2001 (first entry)  
XX  
XX Human secreted protein cDNA sequence #32.  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antifibrotic; hepatocarcinoma; antidiabetic; antiinflammatory; anticancer;  
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiatic; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200058335-A1.

XX  
XX 05-OCT-2000.  
XX  
XX 22-MAR-2000; 2000WO-US007534.  
XX  
XX 26-MAR-1999; 99US-0126598P.  
XX 22-DEC-1999; 99US-0171504P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX  
XX Rosen CA, Ruben SM, Komatsu G;  
XX  
XX WPI; 2000-611702/58.  
XX P-PSDB; AAB37379.  
XX  
XX Nucleic acids encoding human secreted proteins, used to treat, prevent,  
XX ameliorate or diagnose conditions such as cancer, and autoimmune diseases  
XX e.g. arthritis.  
XX  
XX Claim 1; Page 337; 387pp; English.  
XX  
XX The invention relates to the isolation of genes AAC68081-C68127 encoding  
XX 47 human secreted proteins AAB37348-B37394. The genes can be used to  
XX generate fusion proteins by linking to the gene for the human  
XX immunoglobulin G Fc portion (AAC68072) for increasing the stability of  
XX the fusion protein as compared to the human protein only. The genes and  
XX proteins are useful for preventing, ameliorating or treating medical  
XX conditions, e.g. by protein or gene therapy. The genes are isolated from  
XX a range of human tissues disclosed in the specification. The nucleic  
XX acids, proteins, antibodies and (ant)agonists are useful in the  
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
XX immune disorders e.g. Addison's disease, diabetes mellitus, Crohn's disease,  
XX anaplasia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX cardiovascular disorders such as myocardial infarction and ischemias; (d) wound healing  
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX infectious diseases such as viral, bacterial, fungal and parasitic  
XX infections  
XX  
XX Sequence 1470 BP; 386 A; 342 C; 355 G; 382 T; 0 U; 5 Other;  
XX  
XX Query Match 6.9%; Score 34.8; DB 3; Length 1470;  
XX Best Local Similarity 54.8%; Pred. No. 2.8;  
XX Matches 69; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
XX  
XX QY 373 GAGGCTTTGGGGAAGAGAGAGAGGCACTGTTGTGTGTAAGTTTGACTCT 432  
XX DB 71 GAAGCAGTGGGGAAGAGAGAGAGAGGAGTGTGCTGATGTTTAACTCTGACACT 130  
XX QY 433 AACAAATTTGAGACATAGATGACATTTGTCTGACATTAACAACTAGACTCATGTGG 492  
XX DB 131 CACATATAGATTTCACAGTACCCCTGCTGCCCCCAAGACATCCAAAGCAAGATGG 190  
XX QY 493 GTTCTG 498  
XX DB 191 ATGGTG 196  
XX  
XX RESULT 7  
XX ADQ39113/C  
XX ID ADQ39113 standard; DNA; 4352 BP.  
XX  
XX ADQ39113;  
XX  
XX 18-NOV-2004 (first entry)  
XX  
XX Human SNP containing myocardial infarction-associated gene, SEQ ID 776.  
XX  
XX Myocardial infarction; detection; single nucleotide polymorphism; SNP;  
XX  
XX cardiatic; gene therapy; human; gene; de.

XX Homo sapiens.  
 OS  
 XX WO2004058052-A2.  
 FN  
 PD 15-JUL-2004.  
 XX  
 PF 22-DEC-2003; 2003WO-US040978.  
 XX  
 PR 20-DEC-2002; 2002US-0434778P.  
 XX  
 PR 10-MAR-2003; 2003US-0453135P.  
 XX  
 PR 30-APR-2003; 2003US-0466412P.  
 XX  
 PR 23-SEP-2003; 2003US-0504955P.  
 XX  
 PA (APPL-) APPLERA CORP.  
 XX  
 PI Cargill M, Devlin JJ, Iakoubova O;  
 XX  
 DR WPI; 2004-533949/51.  
 XX  
 DR P-PSDB; ADQ39941.  
 XX  
 PT Identifying an individual who has an altered risk for developing  
 PT myocardial infarction by detecting a single nucleotide polymorphism in  
 PT the individual's nucleic acids.  
 XX  
 PS Claim 7; SEQ ID NO 776; 145bp; English.  
 XX  
 CC The invention relates to a novel method for identifying an individual who  
 CC has an altered risk for developing myocardial infarction. The method  
 CC comprises detecting a single nucleotide polymorphism (SNP) in any one of  
 CC the nucleotide sequences given in the specification in the individual's  
 CC nucleic acids, where the presence of the SNP is correlated with an  
 CC altered risk for myocardial infarction in the individual. The invention  
 CC further comprises: an isolated nucleic acid molecule comprising at least  
 CC 8 contiguous nucleotides where one of the nucleotides is an SNP given in  
 CC the specification or its complement and encoding any one of the amino  
 CC acid sequences given in the specification; an isolated polypeptide  
 CC comprising an amino acid sequence given in the specification; an antibody  
 CC that specifically binds to the polypeptide or its antigen-binding  
 CC fragment; an amplified polynucleotide containing an SNP given in the  
 CC specification and which is between about 16 and 1000 nucleotides in  
 CC length; a kit for detecting a SNP in a nucleic acid, comprising the  
 CC polynucleotide, a buffer and an enzyme; a method of detecting a SNP in a  
 CC nucleic acid molecule; a method of detecting a variant polypeptide; and a  
 CC method for identifying an agent useful in treating or preventing  
 CC myocardial infarction. The novel detection method has cardiant activity.  
 CC The nucleic acids of the invention may be used in gene therapy. The  
 CC method is useful in identifying an individual who has an increased or  
 CC decreased risk for developing myocardial infarction and for preparing a  
 CC composition for treating or preventing myocardial infarction. This  
 CC polynucleotide sequence represents a human myocardial infarction-  
 CC associated gene containing one or more SNPs of the invention. Note: This  
 CC sequence was not shown in the specification. The sequence has come from  
 CC an electronic sequence listing downloaded from the WIPO website.  
 XX  
 SQ Sequence 4352 BP; 1108 A; 1051 C; 1209 G; 970 T; 0 U; 14 Other;  
 Query Match 6.9%; Score 34.6; DB 13; Length 4352;  
 Best Local Similarity 53.3%; Pred. No. 5.4;  
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

RESULT 8  
 ABD32629\_4/c  
 Continuation (5 of 6) of ABD32629 from base 400001 (Human cancer-associated genomic DNA  
 WP Sequence split into 6 fragments LOCUS ABD32629 Accession Abd32629  
 WP Fragment Name Begin End  
 WP ABD32629\_0 1 110000  
 WP ABD32629\_1 100001 210000  
 WP ABD32629\_2 200001 310000  
 WP ABD32629\_3 300001 410000  
 WP ABD32629\_4 400001 510000  
 WP ABD32629\_5 500001 518360  
 Query Match 6.9%; Score 34.4; DB 13; Length 110000;  
 Best Local Similarity 52.9%; Pred. No. 28;  
 Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 12 ACATTCTTCATTGATTCATCTTGATCTTCTTTCTTGAACAATACGTAACCTGACCTTTG 71  
 DB 80461 AGATTTTGAATTCAGACTCTAAGACTTCTCTCCAGATCACAATGCACTGATACCAT 80402

QY 72 TTGGCAGCTTCGAATTTGACACATAGGCGTTGAACTTGAGCAAGATCCTCATCCTGTC 131  
 DB 80401 TTCTGATCTGATCTTGAAAGGTGGGCTTAAACTGATGACGATGATGATAGTGC 80342

QY 132 ACCCAGCATCTGCGTAC 151  
 DB 80341 AGCAAAACCATGCGACAC 80322

RESULT 9  
 ADT05514/c  
 ID ADT05514 standard; DNA; 7991 BP.  
 AC ADT05514;  
 DT 02-DEC-2004 (first entry)  
 XX  
 DE Haemophilus influenzae (NTHI) contig DNA sequence - SEQ ID 550.  
 XX  
 KW middle ear bacterial infection; nasopharynx bacterial infection; ds;  
 KW contig.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO2004078949-A2.  
 XX  
 PD 16-SEP-2004.  
 XX  
 PD 05-MAR-2004; 2004WO-US007001.  
 XX  
 PF 06-MAR-2003; 2003US-0453134P.  
 XX  
 PR (CHIL-) CHILDRENS HOSPITAL INC.  
 XX  
 PA Bakaletz LO, Munson RS, Dyer DW;  
 XX  
 PI WPI; 2004-662422/64.  
 XX  
 DR New polynucleotides of non-typeable strain of Haemophilus influenzae,  
 XX useful for treating or preventing NTHI bacterial infections of the middle  
 PT ear and/or nasopharynx.  
 PT  
 PS Example 1; SEQ ID NO 550; 88bp; English.  
 XX  
 CC The invention comprises nucleotide sequences (genes) from the genome of a  
 CC non-typeable strain of Haemophilus influenzae (NTHI). The NTHI DNA  
 CC sequences of the invention are useful for treating or preventing NTHI  
 CC bacterial infections of the middle ear and/or nasopharynx. The present  
 CC nucleic acid represents an NTHI contig sequence of the invention.  
 XX  
 SQ Sequence 7991 BP; 2442 A; 1657 C; 1486 G; 2405 T; 0 U; 1 Other;

Query Match 6.8%; Score 34.2; DB 13; Length 7991;  
Best Local Similarity 53.3%; Pred. No. 9.7;  
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

216 GAAATAGGCGCTGATCTCATCTCGGACACATATTAATGACCTAGATCCAGAG 275  
2844 GCATTATATTTCTTTTGTGATGTAATTTCAACAAACCTTAATCCATCCCTGGG 2785

276 ATCTGTCAGAGAAACAAAGATCTCTCTGTCGATCTTTCTTCAAGCAGTAGTGTG 335  
2784 ATGCAGAAATATAAATGGATATTGCTGCTTACCTTTACTTAACACTATTATTATG 2725

336 ACTCCACTAACTGA 350  
2724 GTTTTACGCTATTTA 2710

RESULT 10  
ADQ97152/c  
ID ADQ97152 standard; DNA; 58651 BP.  
AC ADQ97152;  
XX 07-OCT-2004 (first entry)  
XX Human cancer associated sequence HD08-008, SEQ ID 128.  
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.  
XX Homo sapiens.  
XX WO2004060304-A2.  
XX 22-JUL-2004.  
XX 22-DEC-2003; 2003WO-US041389.  
XX 27-DEC-2002; 2002US-00330773.  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Malandro MS;  
XX WPI; 2004-543781/52.  
XX New isolated cancer associated nucleic acids comprising at least 10  
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
XX cancers such as leukemia and lymphoma.  
XX Claim 1; SEQ ID NO 128; 199pp; English.  
XX The present invention relates to cancer associated sequences (ADQ97025-  
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
XX data for this patent did not form part of the printed specification, but  
XX CC was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 58651 BP; 15392 A; 12241 C; 13719 G; 17299 T; 0 U; 0 Other;

Query Match 6.8%; Score 34; DB 12; Length 58651;  
Best Local Similarity 54.9%; Pred. No. 29;  
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

380 TGGGGCAAGACGAGGAAAGCACTGTTGTGTGTAAGTTTGACTTAACAAT 439  
58574 TGCAGCCTTAAGTGAAGGCCGTGGCACTGAGGTGAGAGGAGATGGCACTGACAGA 58515

440 TTGAAGCATATGATGATGTTGTCTCAACTAACAACAACCTAAGCTAGTGGGTTCTGT 499  
58514 GAGTAGACAGAGCACTGAGGCTATCTAGAAACAAACAGTACATCAGCGATCAGT 58455

500 TT 501

DB 58454 TT 58453

RESULT 11  
ADQ97379/c  
ID ADQ97379 standard; DNA; 23473 BP.  
XX ADQ97379;  
XX 07-OCT-2004 (first entry)  
XX Human cancer associated sequence HD08-040, SEQ ID 356.  
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.  
XX Homo sapiens.  
XX WO2004060304-A2.  
XX 22-JUL-2004.  
XX 22-DEC-2003; 2003WO-US041389.  
XX 27-DEC-2002; 2002US-00330773.  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Malandro MS;  
XX WPI; 2004-543781/52.  
XX New isolated cancer associated nucleic acids comprising at least 10  
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
XX cancers such as leukemia and lymphoma.  
XX Claim 1; SEQ ID NO 356; 199pp; English.  
XX The present invention relates to cancer associated sequences (ADQ97025-  
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
XX data for this patent did not form part of the printed specification, but  
XX CC was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 23473 BP; 5653 A; 5992 C; 6186 G; 5642 T; 0 U; 0 Other;

Query Match 6.7%; Score 33.8; DB 12; Length 23473;  
Best Local Similarity 48.2%; Pred. No. 22;  
Matches 95; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

283 AGAGAAACAGAGATCTCTGTCGTCATCTTTCTTCAAGCAGTAGTGAATCTCAC 342  
22832 ATATATATATATATATGTCCTTCAAGTCTTTCTTCAACACAGCATCCTGTCTGTG 22773

343 TAAACTGAGTCTCCATCTGAGATCCACAGAGAGCTTTGGGCAAGAGAGGAGG 402  
22772 ACCTAGAGATGACATCTGACACCTTCACAGACTGCTGTGGGTGGAGAGGAGG 22713

403 CACTGTTGTGTTGGTAAAGTTTGACTTAACAATTTGAAGACATAGATGACATTTGTG 462  
22712 GAATGAGGCTATTGTGTAAGAGAGCTTGACCGCTGTGGGATAGCTGCTCATATGTC 22653

463 TCAGACTTAACAACAACC 479  
22652 ACTGCCCACTTCTTC 22636

RESULT 12  
ACN44068/c  
ID ACN44068 standard; DNA; 295096 BP.  
XX ACN44068;





PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-014187P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142820P.  
PR 12-JUL-1999; 99US-0143277P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.

PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 18-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 26-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.6%; Score 33.2; DB 3; Length 1089;  
Best Local Similarity 61.6%; Pred. No. 8.1;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 337 CTCACCTAACTGATCCATCTGAGAGTCACAGAGGCTTTGGGCAAGACAGAG 396  
DB 220 CTTACACACCTTGACGACTTTGTTGAGGACACAGAGGCTTTCCCAATAAGCAGAG 161  
QY 397 GGAAGGCACTGTTGTTGTTGTAAG 422  
DB 160 AGGGTCACGTTGGCGTTGTAAG 135

RESULT 14  
AAC51477/c  
ID AAC51477 standard; DNA; 1417 BP.

AC AAC51477;  
XX  
XX 18-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68668.  
XX

KM Hybridisation assay; Genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway; metabolic pathway;  
KM promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EPI033405-A2.  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129645P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134270P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135333P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151386P.

PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155658P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158332P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.6%; Score 33.2; DB 3; Length 1417;

Best Local Similarity 61.6%; Pred. No. 9.2; Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 337 CTCACAACTGATGCCATCTGAGATGCACAGAGGCTTGGGCAAGACAGAG 396  
DB 358 CTCACAACTTCACTGATCTTGTGAGAACACAAAGGCGTTTCCCAATAAGCAGAG 299  
QY 397 GGAAGCACTGTTGTGTGTAAAG 422  
DB 298 AGGCGTCACTTGGCGCTGTAGAG 273

RESULT 15

AAC46837/c  
ID AAC46837 standard; DNA; 1431 BP.

XX AAC46837;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51586.

XX Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.

XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 200SEP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 02-JUL-1999; 99US-0142154P.  
 PR 03-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144848P.  
 PR 20-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 22-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 23-JUL-1999; 99US-0145276P.  
 PR 26-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145919P.  
 PR 27-JUL-1999; 99US-0145919P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 23-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 30-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.

PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 26-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161939P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 6.6%; Score 33.2; DB 3; Length 1431;  
 Best Local Similarity 61.6%; Pred. No. 9.2;  
 Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 337 CTCACATAACTGAGTTCCTCCATCTGAGATCCACAGAGGCTTTGGGCGACAGACAGAG 396  
 Db 373 CTTCAACACCTTCAGTGAAGCTTTGTTGAGAGACACAGAGGCGTTTCCCAATAGCAGAG 314  
 QY 397 GGAAGGCACTGTTTGTGTTGTTGAAG 422  
 Db 313 AGGCGTACCTTTGGCGTTGTAGAG 288

Search completed: February 10, 2005, 09:41:10  
 Job time : 348.632 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58, Search time 103.92 Seconds  
(without alignments)  
7888.487 Million cell updates/sec

Title: US-09-926-375B-7\_COPY\_11500\_12000

Perfect score: 501  
Sequence: 1 aaccacaacgaacatcttc.....gactcatcgggtctctgtt 501

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35.6	7.1	152466	4 US-09-949-016-12869	Sequence 12869, A
C 2	34.4	6.9	1380890	4 US-09-949-016-14720	Sequence 14720, A
C 3	34.2	6.8	144362	4 US-09-949-016-16066	Sequence 16066, A
4	33.8	6.7	1055	4 US-09-806-7088-23	Sequence 23, Appl
5	33.4	6.7	101674	4 US-09-949-016-12033	Sequence 12033, A
6	33	6.6	27250	4 US-09-949-016-15537	Sequence 15537, A
C 7	32.6	6.5	601	4 US-09-949-016-64253	Sequence 64253, A
8	32.6	6.5	183202	4 US-09-949-016-13614	Sequence 13614, A
9	32.4	6.5	832	4 US-09-621-976-2813	Sequence 2813, Ap
C 10	32	6.4	91665	4 US-09-949-016-12234	Sequence 12234, A
11	32	6.4	134890	4 US-09-949-016-15602	Sequence 15602, A
12	31.4	6.3	786431	4 US-09-751-389-3	Sequence 3, Appl
13	31.2	6.2	222452	4 US-09-949-016-12968	Sequence 12968, A
C 14	31	6.2	2109	4 US-09-949-016-5757	Sequence 5757, Ap
C 15	31	6.2	26011	4 US-09-949-016-17499	Sequence 17499, A
16	31	6.2	94133	4 US-09-949-016-11901	Sequence 11901, A
17	31	6.2	94133	4 US-09-949-016-12713	Sequence 12713, A
18	31	6.2	94135	4 US-09-949-016-15934	Sequence 15934, A
19	31	6.2	94135	4 US-09-949-016-15935	Sequence 15935, A
20	31	6.2	94135	4 US-09-949-016-15936	Sequence 15936, A
21	31	6.2	94135	4 US-09-949-016-15937	Sequence 15937, A
22	30.8	6.1	601	4 US-09-949-016-178097	Sequence 178097, A
C 23	30.8	6.1	98302	4 US-09-328-475C-19	Sequence 19, Appl
C 24	30.6	6.1	980	4 US-09-949-016-15250	Sequence 15250, A
C 25	30.6	6.1	21679	4 US-09-949-016-13356	Sequence 13356, A
C 26	30.6	6.1	31820	4 US-09-949-016-13356	Sequence 13356, A
C 27	30.6	6.1	232024	4 US-09-949-016-13477	Sequence 13477, A

C 28	30.6	6.1	784019	4 US-09-949-016-14033	Sequence 14033, A
C 29	30.6	6.1	828152	4 US-09-949-016-12777	Sequence 12777, A
C 30	30.4	6.1	402	4 US-09-583-110-842	Sequence 842, App
31	30.4	6.1	38368	4 US-09-949-016-12858	Sequence 12858, A
32	30.2	6.0	256	1 US-08-616-368A-25	Sequence 25, Appl
33	30.2	6.0	256	3 US-09-054-298-25	Sequence 25, Appl
34	30.2	6.0	256	3 US-08-818-655-25	Sequence 25, Appl
35	30.2	6.0	256	4 US-09-305-839-25	Sequence 25, Appl
C 36	30.2	6.0	601	4 US-09-949-016-55292	Sequence 55292, A
37	30.2	6.0	107329	4 US-09-949-016-12663	Sequence 12663, A
38	30.2	6.0	107329	4 US-09-949-016-12664	Sequence 12664, A
39	30.2	6.0	107330	4 US-09-949-016-15408	Sequence 15408, A
40	30.2	6.0	107330	4 US-09-949-016-15409	Sequence 15409, A
41	30.2	6.0	107330	4 US-09-949-016-15410	Sequence 15410, A
42	30.2	6.0	107330	4 US-09-949-016-15411	Sequence 15411, A
43	30.2	6.0	107750	4 US-09-949-016-12662	Sequence 12662, A
44	30.2	6.0	107751	4 US-09-949-016-15412	Sequence 15412, A
45	30.2	6.0	107751	4 US-09-949-016-15413	Sequence 15413, A

## ALIGNMENTS

```

RESULT 1
US-09-949-016-12869/C
; Sequence 12869, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12869
; LENGTH: 152486
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12869
Query Match      7.1% Score 35.6; DB 4; Length 152486;
Best Local Similarity 62.2%; Pred. No. 3;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY      23 TTGATTCATTCGATCTTCTTTAGACAACTGACCTGACCTGTTGGACGTC 82
      113516 TTTTCATCTTCGTATTCCTTCCTTCACATCTTCACCTGCGCCCTGTTCTTCTT 113457
Db      113456 CTGTTGCTAAAGGGGCACTCTCCTGG 113427

RESULT 2
US-09-949-016-14720/C
; Sequence 14720, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```

; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14720  
 ; LENGTH: 390890  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(390890)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-14720

Query Match  
 Best Local Similarity 52.9%; Score 34.4; DB 4; Length 390890;  
 Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 12 ACATTCTTCATTGATTCATTCGATCTCTTTCTAGACAATGTAAGCCCTTG 71  
 DB 364776 AGATTCTTATTCAGACTCTTAAGACTTCTCTCAGATTCACATGCACTGATACACT 364717  
 QY 72 TTGGCAGTCTCAAGTTGAACAATAGGCGCTTGAACCTTGGCAAGGTCACTGTC 131  
 DB 364716 TTGGTATCTGATCTTAAGAAAGGTGGGCTTAAACCTAGATGACGATGATGATGTC 364657  
 QY 132 ACCCAAGCATCCGCGGTGAC 151  
 DB 364656 AGCAACACACCATGACGACAC 364637

RESULT 3  
 US-09-949-016-16066/c  
 ; Sequence 16066, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16066  
 ; LENGTH: 144362  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(144362)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-16066

Query Match  
 Best Local Similarity 52.4%; Score 34.2; DB 4; Length 144362;  
 Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 2 ACCCAAGACAATTTCTTCAATTCATTCGATCTCTTTCTAGACAATGTAAG 61  
 DB 27696 AGACCCCTTACTTACTTACACGTAATCTGTTGGCTCTACATGAACTAACCAAT 27637  
 QY 62 TGACCCCTTGTGGCAGTCTCAAGTTGACAACTAGAGGCTTTGAACTTGGCAAGGTC 121

DB 27636 TCAACCACTTCTCTCAAACTCACTGTATACCACTTGTCTAGACACCTTCATCCCTC 27577  
 QY 122 CATCATGTCAACCAAGATCCT 144  
 DB 27576 CTAAATTAACAACAAGCCCTCT 27554

RESULT 4  
 US-09-806-708B-23  
 ; Sequence 23, Application US/09806708B  
 ; Patent No. 6784342  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The University of British Columbia  
 ; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
 ; FILE REFERENCE: 4810-58741  
 ; CURRENT APPLICATION NUMBER: US/09/806,708B  
 ; PRIOR FILING DATE: 2001-04-03  
 ; PRIOR APPLICATION NUMBER: US 60/147,133  
 ; PRIOR FILING DATE: 1999-08-04  
 ; NUMBER OF SEQ ID NOS: 23  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 23  
 ; LENGTH: 1055  
 ; TYPE: DNA  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; NAME/KEY: promoter  
 ; LOCATION: (1)...(1055)  
 ; OTHER INFORMATION: consensus sequence of A.t. and L.a. PAB1 promoters  
 US-09-806-708B-23

Query Match  
 Best Local Similarity 26.0%; Score 33.8; DB 4; Length 1055;  
 Matches 47; Conservative 47; Mismatches 87; Indels 0; Gaps 0;

QY 1 AACCCAGACAATCTCTCATTCATTCGATCTCTTCTTCTAGACAATGTAAG 60  
 DB 303 AAMVAAAMMMWSATTTWMAAAMTSRTWTYYGKRTAMNNNGTTCWRMAWTKMMK 362  
 QY 61 CTGACCCCTTGTGGCAGTCTCAAGTTGACAACATGAGGCTTGAACCTTGGCAAGGT 120  
 DB 363 TGGTTNNNGRTYGTGTTKMATTTAKANNCTTAAMKKTCTMNTTAAKATTTWAT 422  
 QY 121 CCATCAGTGCACCAAGCATCTCGGTGACCTTTGGGTGAATATCTGGCAACCTT 180  
 DB 423 CTXKMTGTSYRYAARATYAWMTTRRYAANNNTTAKACGWTYYKCCCTAANTA 482  
 QY 181 A 181  
 DB 483 A 483

RESULT 5  
 US-09-949-016-12033  
 ; Sequence 12033, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12033  
 ; LENGTH: 101674

TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)... (101674)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12033

Query Match  
Best Local Similarity 58.6%; Pred. No. 13;  
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 307 TGCATCTCTTCAAGCAGTGTGATCTCACTAACTAGTCCCTGAGCT 366  
DB 77931 TTCTTTTATTATACAGCAGCAGCTATATCCACACTGATGAGTTCTTTGACAG 77990  
QY 367 CCACGAGGCTTTGGGCAAGAGAGAGGAGGAC 405  
DB 77991 CCATTAAAGAGAGGAGGAGTGAAGAGGCTGAAACC 78029

RESULT 6  
US-09-949-016-15537  
Sequence 15537, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15537  
LENGTH: 27250  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15537

Query Match  
Best Local Similarity 53.5%; Pred. No. 9;  
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 125 CACTGTCAACCAAGCATCTGGTGACCTTGGGTTGGAATCTTGCTAACCTTAGAT 184  
DB 25818 CACAGTCTCCCAAGTACTGACTACAGGTGTCTACACATCCCTGTAATTTTAG 25877  
QY 185 ATTTTCTTGGAGTCTTTAGAACATCCAGAAATAGGCTGATTTCTACTCTGGAC 244  
DB 25878 TTTTGTATTATATTATTATTATTATTAGAAACAGATCTCCTGTCAACCCAGGC 25937  
QY 245 CACATATA 253  
DB 25938 TAGAGTACA 25946

RESULT 7  
US-09-949-016-64253/C  
Sequence 64253, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 64253  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-64253

Query Match  
Best Local Similarity 49.7%; Pred. No. 1,6;  
Matches 83; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 228 GATTCATCTCTGGGACCAATATATGACCTTAGATCCAGAGATCGTCAGAGA 287  
DB 227 GACCATCAAAATGAGAAAAAACAATATATACAGACTTACGTGATTATTCAGAGA 168  
QY 288 AACAGATCTCTCTGTGTGATCTCTTCAAGCAGTGTGATGATCTCACTAAC 347  
DB 167 ATTAATATGCAAAATGCTTGTGTTTAAAAATGCAAGGAAAGTAACTGTACTAAA 108  
QY 348 TGAGTCCCATCTGAGATCCACAGAGGCTTTGGGCAAGACAG 394  
DB 107 TGAGACACTTTATATAGAGAGAGAAAGATTGGCGTAAGATAG 61

RESULT 8  
US-09-949-016-13614  
Sequence 13614, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13614  
LENGTH: 183202  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)... (183202)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13614

Query Match  
Best Local Similarity 49.7%; Pred. No. 33;  
Matches 83; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 228 GATTCATCTCTGGGACCAATATATGACCTTAGATCCAGAGATCGTCAGAGA 287  
DB 11606 GACCATCAAAATGAGAAAAAACAATATATACAGACTTACGTGATTATTCAGAGA 116065  
QY 288 AACAGATCTCTCTGTGTGATCTCTTCAAGCAGTGTGATGATCTCACTAAC 347  
DB 11606 ATTAATATGCAAAATGCTTGTGTTTAAAAATGCAAGGAAAGTAACTGTACTAAA 116125

Qy 348 TGAGTTCCTGAGAGTCCAGAGAGGCTTTGGGGCAAGAGCAG 394  
116126 TGAGGACACTTTATGAGAGAGAGAAAGATTGGCGTAAGATGAG 116172

RESULT 9  
US-09-621-976-2813  
; Sequence 2813, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 2813  
; LENGTH: 832  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 235..399  
US-09-621-976-2813

Query Match 6.5%; Score 32.4; DB 4; Length 832;  
Best Local Similarity 9.9%; Pred. No. 2.2; Indels 1; Gaps 1;  
Matches 35; Conservative 166; Mismatches 152;

Qy 86 TTGACAACATAGAGCTTTGAACTTGACAAGATCATCTGACCCAGCATCTG 145  
10 TWAKCMTKKMSKMSYMWKMYMKTKWMRKMKKKKMKWMTWYMYRMAWMTYK 69  
Qy 146 GGTGACCTTTGGTGAATATCTGGCTACTT-TAGATATTTCTTTGAGATCTT 204  
70 KKAMCRTKTKKKKKKGGMMWYMGWRSYMAWTRTWGAYRSMWYMRHKMKKKAY 129  
Qy 205 AGAACAATCCAGAAATAGGCTTGATCTCATCTGAGACCAATATAAGTCACTG 264  
130 YRKTTCYSSKGTWTKMKRKAATWTKYMAAIRYMMWMTKRMASWYCWMMWK 189  
Db 265 AATCCAGAGATCGTGAGAGAAAGAGATCTCTCTGTCATCTTTCAAGC 324  
190 ARKMTWRKRSYASASARCCYSCWGMASWKMWRMRGMAWGATGACMKAMRASCM 249  
Qy 325 AGTGAGTAGTACTCCCACTGAAGTCCCATCTGAGAGTCCACAGAGAGGCTTGGGG 384  
250 RRRYAGSKTSYSMMWCTRSWYCYTKARMTGYCYRKGWKGKRGWYASKTKYMKR 309  
Db 385 CAAGAAGCAGAGGAGGACCTGTTGTGTTGTAAGTTTGACTCTTAACAA 438  
310 WMMCMAMMYSTCTRASMMWRMYTMMKMKMYAABAAMWMAWMAWMBRA 363

RESULT 10  
US-09-949-016-12234/c  
; Sequence 12234, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12234  
; LENGTH: 91665  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12234

Query Match 6.4%; Score 32; DB 4; Length 91665;  
Best Local Similarity 45.3%; Pred. No. 37;  
Matches 116; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 60 ACTGACCCCTTTGGGAGCTCTCAAGTTTGACAATAGAGGCTTTGAATGCGACAGG 119  
9823 AGTAATGAGTGTGACTGACGAGCCCAAGAGGTATGTATATTTGCAATTTGAATAT 9764  
Db 120 TCCATCACTGTCAACCAAGCATCTGGGTGACCTTTGGGTGAATATCTTGCTAACT 179  
Qy 9763 ATCTTAATGAGAGCAAGAAATTTAGTAGATATCTCTTGAGCAGCTCTGTTAACT 9704  
Db 180 TAGATATTTCTTTGGAGTATCTTTGAATCCAGAGAAATAGGCTTTGATCTCATCCT 239  
Qy 9703 CAGAGAGAACTAAAGCCGCTGTCTAAATATGAGTTGTATATGACAGGTAAGAAAT 9644  
Db 240 GGCACCAATATATAGTCAACCTGAATCCAGAGATCGTGACAGAAACAGAGATCTC 299  
Qy 9643 GCACCCCAAAATATGCACTTTGGAATTTAGAAATATGAGAAACAGAGAAAGTTTCTC 9584  
Db 300 TCTCGTGTGATCTCTT 315  
Qy 9583 TGACCTTTGCTCCTT 9568

RESULT 11  
US-09-949-016-15602  
; Sequence 15602, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15602  
; LENGTH: 134890  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15602

Query Match 6.4%; Score 32; DB 4; Length 134890;  
Best Local Similarity 60.2%; Pred. No. 45;  
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 84 AGTTGACAACATAGAGCTTTGAACCTTGACAGAGGCTCATGTCATCCCAAGCATCC 143  
Db 89981 ATTTATGAAAAGAGATTTGGAGCTGGGCAACATGGCTCATGCTATATCCAGACTT 90040  
Qy 144 TGGGTGACCTTTGGGTTGGAATATCTTG 171  
Db 90041 TGGAGGCCAAGAGAGATGATCATCTTG 90068

RESULT 12





ORGANISM: Human  
US-09-949-016-17499

Query Match 6.2%; Score 31; DB 4; Length 26011;  
Best Local Similarity 59.8%; Pred. No. 41;  
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy	260	CCTAGATCCGAGAGATCGTGCAAGAAACAAGATCTCTCGTGCACTCTTC	319
Db	23278	CATAGAGCACATGAAGAACAGCAGTGAAACCAAGATTCTAGCAAGGCCAGCTATTAGC	23219
Qy	320	AAAGCAGTGAGTAGTCACTAA	346
Db	23218	AAAGCAGTAAGCAGAACTGACTAGA	23192

Search completed: February 10, 2005, 17:42:54  
Job time : 108.92 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:13:58 ; Search time 365.063 Seconds  
(without alignments)  
8097.594 Million cell updates/sec

Title: US-09-926-375b-7\_COPY\_11500\_12000

Perfect score: 501  
Sequence: 1 aaccacacgaacatctctc.....gactcatgtgggtctcttt 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5378673 seqs, 295022984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	7.4	3555	US-10-282-122A-16204	Sequence 16204, A
2	36.8	7.3	49088	US-10-087-192-13	Sequence 13, App1
3	36.8	7.3	56773	US-10-311-053-47	Sequence 47, App1
4	36	7.2	503	US-10-027-632-138049	Sequence 138049, A
5	36	7.2	503	US-10-027-632-138049	Sequence 138049, A
6	36	7.2	1032	US-10-424-599-132734	Sequence 132734, A
7	35.4	7.1	34511	US-10-085-117-307	Sequence 307, App
8	34.6	6.9	51806	US-10-719-993-6892	Sequence 6892, App
9	34.6	6.9	4352	US-10-741-600-776	Sequence 776, App
10	34.4	6.9	634	US-10-424-599-80920	Sequence 80920, A
11	34.4	6.9	4881	US-10-437-963-40327	Sequence 40327, A

C 12	34.4	6.9	518360	US-10-367-094-125	Sequence 125, App
C 13	34.2	6.8	2859	US-10-027-632-264212	Sequence 264212, A
C 14	34.2	6.8	2859	US-10-027-632-264212	Sequence 264212, A
C 15	33.6	6.7	295096	US-10-087-192-331	Sequence 331, App
C 16	33.4	6.7	1536	US-10-424-599-120851	Sequence 120851, A
C 17	32.8	6.5	599	US-10-027-632-280124	Sequence 280124, A
C 18	32.8	6.5	599	US-10-027-632-280124	Sequence 280124, A
C 19	32.6	6.5	311	US-09-864-408A-6439	Sequence 6439, App
C 20	32.6	6.5	321	US-10-674-124A-2052	Sequence 2052, A
C 21	32.6	6.5	508	US-10-027-632-141634	Sequence 141634, A
C 22	32.6	6.5	508	US-10-027-632-141634	Sequence 141634, A
C 23	32.6	6.5	89829	US-10-322-281-359	Sequence 359, App
C 24	32.6	6.5	197526	US-10-322-281-359	Sequence 359, App
C 25	32.4	6.5	1462	US-10-424-599-87448	Sequence 87448, A
C 26	32.4	6.5	1793	US-10-425-114-29850	Sequence 29850, A
C 27	32.4	6.5	1816	US-10-424-599-18143	Sequence 18143, A
C 28	32.2	6.4	315	US-09-783-590-1719	Sequence 1719, App
C 29	32.2	6.4	37925	US-09-984-429-701	Sequence 701, App
C 30	32.2	6.4	211257	US-10-087-192-529	Sequence 529, App
C 31	32.2	6.4	1808	US-10-369-493-33341	Sequence 33341, A
C 32	31.8	6.3	434	US-10-027-632-38203	Sequence 38203, A
C 33	31.8	6.3	434	US-10-027-632-38203	Sequence 38203, A
C 34	31.8	6.3	542	US-10-027-632-63825	Sequence 63825, A
C 35	31.8	6.3	542	US-10-027-632-65209	Sequence 65209, A
C 36	31.8	6.3	542	US-10-027-632-296771	Sequence 296771, A
C 37	31.8	6.3	542	US-10-027-632-63825	Sequence 63825, A
C 38	31.8	6.3	542	US-10-027-632-65209	Sequence 65209, A
C 39	31.8	6.3	542	US-10-027-632-296771	Sequence 296771, A
C 40	31.8	6.3	825	US-10-282-122A-16747	Sequence 16747, A
C 41	31.8	6.3	3094	US-10-027-632-177175	Sequence 177175, A
C 42	31.8	6.3	3094	US-10-027-632-177175	Sequence 177175, A
C 43	31.8	6.3	5969	US-10-172-086-45	Sequence 45, App1
C 44	31.8	6.3	5969	US-10-240-585C-135	Sequence 135, App
C 45	31.8	6.3	5969	US-10-311-507-83	Sequence 83, App1

#### ALIGNMENTS

RESULT 1  
US-10-282-122A-16204  
; Sequence 16204, Application US/10282122A  
; Publication No. US20040029129A1  
GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlson, Karl  
; APPLICANT: Zyckind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
TITLE OR INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EPI7RA.034A  
CURRENT APPLICATION NUMBER: US/10/282.122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625

```
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO: 16204
/ LENGTH: 3555
/ TYPE: DNA
/ ORGANISM: Clostridium botulinum
US-10-282-122A-16204
```

```
Query Match
Best Local Similarity 56.0%; Pred. No. 1.2; DB 17; Length 3555;
Matches 70; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
```

```
QY 181 AGATATTTCCTTGGAGTATCTTTAGACATCCAGGAATAGGCTTGATTCATCTCG 240
DB 1947 AGATATAATAGCAGATAGTTATATAGACAGCCCTATGGCTGAATACCAAGATA 2006
QY 241 GGACCAATATATAGTACCCCTTGAATCCAGAGATCGTGAGAGAAACAAGATCTCT 300
DB 2007 TATAGCTAATAAATAATACCTTTAATAATAGAAATCAACAGAAAACTAAGACTTAT 2066
QY 301 CTGCT 305
DB 2067 ATTGT 2071
```

## RESULT 2

```
US-10-087-192-13/C
/ Sequence 13, Application US/10087192
/ Publication No. US20020182586A1
/ GENERAL INFORMATION:
/ APPLICANT: Morris, David W.
/ APPLICANT: Engelhard, Eric K.
/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
/ TITLE OF INVENTION: CANCER
/ FILE REFERENCE: 529452000122
/ CURRENT APPLICATION NUMBER: US/10/087,192
/ CURRENT FILING DATE: 2002-03-01
/ PRIOR APPLICATION NUMBER: US 09/747,377
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 09/798,586
/ PRIOR FILING DATE: 2001-03-02
/ NUMBER OF SEQ ID NOS: 2059
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13
/ LENGTH: 49088
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(49088)
/ OTHER INFORMATION: n = A,T,C or G
US-10-087-192-13
```

```
Query Match
Best Local Similarity 7.3%; Score 36.8; DB 13; Length 49088;
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 255 GTCAACCTAGATCCAGAGATGTCGAGAGAAACAAGATCTCTCGTGTGCATCTT 314
DB 45517 GTCTCAAAACAATCTGTGGAGTTGACCAAGTAAGATTAAGTTTAAACCCCTGAACT 45458
QY 315 TCTTCAAGCAGTAGTAGTACCTCCACTAACTGAAGTTCCCAT 358
DB 45457 ACATATAAGCAGAAAGAAATGACCCCACTAAGTAGTCTTTTCAT 45414
```

```
RESULT 3
US-10-331-053-47/C
/ Sequence 47, Application US/10331053
/ Publication No. US2004019778A1
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc S. Malandro
/ TITLE OF INVENTION: Novel Compositions and Methods in Cancer
/ FILE REFERENCE: 529452001100
/ CURRENT APPLICATION NUMBER: US/10/331,053
/ CURRENT FILING DATE: 2002-12-26
/ NUMBER OF SEQ ID NOS: 86
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 47
/ LENGTH: 56773
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(56773)
/ OTHER INFORMATION: n = A,T,C or G
US-10-331-053-47
```

```
Query Match
Best Local Similarity 59.6%; Pred. No. 5.5; DB 18; Length 56773;
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 255 GTCAACCTAGATCCAGAGATGTCGAGAGAAACAAGATCTCTCGTGTGCATCTT 314
DB 48995 GTCTCAAAACAATGCTGGAGTTGACCAAGTAAGATTAAGTTTAAACCCCTGAACT 48936
QY 315 TCTTCAAGCAGTAGTAGTACCTCCACTAACTGAAGTTCCCAT 358
DB 48935 ACATATAAGCAGAAAGAAATGACCCCACTAAGTAGTCTTTTCAT 48892
```

## RESULT 4

```
US-10-027-632-138049
/ Sequence 138049, Application US/10027632
/ Publication No. US20020198371A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827,129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 138049
/ LENGTH: 503
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-138049
```

```
Query Match
Best Local Similarity 7.2%; Score 36; DB 13; Length 503;
Matches 105; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
```

```
QY 185 ATTTCTTTGAGATCTTTAGAACATCCAGAAATAGGGCTTGATTTCTCATCTGGGAC 244
    |||||
DB 185 ACTGCTTTGGGGGGCTTATGACCTGTAGAGAGAGGGGGCTGAGGACAGCCCATGAT 244
QY 245 CACAAATATAGTACACCTTGAATCCACGAGAGATCGTCAAGAGAAACAAGATCTCTCG 304
    |||||
DB 245 TTTCTCACACTCTCTTGAAGTCTACCTCAAGGGTGTCTTCTCATACAGCTCTCTCT 304
QY 305 TGTGACCTCTTCTTCAAGCAGTGAAGTGAAGTCTCACTAACTGAGTTCCATCTGAGA 364
    |||||
DB 305 GTTCTCTCTCTTGAAGTGTGCCCCCTCTCTGCTGACGAGAACTGCTCTTTGTAC 364
QY 365 GTCCACAGAGGCTTTGGGCAAGAGAGAGGAGGCA 404
    |||||
DB 365 ATCTCAGGCTGAATGAGAGACAGAGAGGTCAGAGGGA 404

RESULT 5
US-10-027-632-138049
; Sequence 138049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138049
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138049

Query Match
Best Local Similarity 47.7%; Pred. No. 0.92; Length 503;
Matches 105; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 185 ATTTCTTTGAGATCTTTAGAACATCCAGAAATAGGGCTTGATTTCTCATCTGGGAC 244
    |||||
DB 185 ACTGCTTTGGGGGGCTTATGACCTGTAGAGAGAGGGGGCTGAGGACAGCCCATGAT 244
QY 245 CACAAATATAGTACACCTTGAATCCACGAGAGATCGTCAAGAGAAACAAGATCTCTCG 304
    |||||
DB 245 TTTCTCACACTCTCTTGAAGTCTACCTCAAGGGTGTCTTCTCATACAGCTCTCTCT 304
QY 305 TGTGACCTCTTCTTCAAGCAGTGAAGTGAAGTCTCACTAACTGAGTTCCATCTGAGA 364
    |||||
DB 305 GTTCTCTCTCTTGAAGTGTGCCCCCTCTCTGCTGACGAGAACTGCTCTTTGTAC 364
QY 365 GTCCACAGAGGCTTTGGGCAAGAGAGAGGAGGCA 404
    |||||
DB 365 ATCTCAGGCTGAATGAGAGACAGAGAGGTCAGAGGGA 404

RESULT 6
```

```
US-10-424-599-132734/c
; Sequence 132734, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 132734
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90866C.1
US-10-424-599-132734

Query Match
Best Local Similarity 55.6%; Pred. No. 1.3; Length 1032;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 112 GCACAGTTCATCACTGTACCCCAAGATCTCTGGGTGACCTTGGGTTGAATATCTTG 171
    |||||
DB 647 GCACAGTTTAAATCCCTATATGAGCTCTCTCTTTCATTTATTTGGTTCTCTT 588
QY 172 GCTAACCTTAGATATTTCTTTGAGATCTTTAGAACATCCAGAAATAGGCTTGAT 231
    |||||
DB 587 TCAAACACAGATAGTCTATGCTTCCCTCCCTACAACTCAAGAAATATTCATCAT 528
QY 232 CTCA 235
    |||
DB 527 CTTA 524

RESULT 7
US-10-085-117-307/c
; Sequence 307, Application US/10085117
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; PRIOR FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 307
; LENGTH: 34511
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(34511)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-307

Query Match
Best Local Similarity 7.1%; Score 35.4; DB 17; Length 34511;
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 272 GGAGATGTCAGAGAAACAAGATCTCTCTGTGTCATCTTTCTTCAAGAGAGAGT 331
    |||||
DB 31333 GGAGATGTCAGAGAGAGTCAACATCTCTGTCAATTTCAATTAATTTTCGAAGCTATGTT 31274
QY 332 AGTGAATCCACTAACTGAGTCTCCATCTGAGATCCACAGAGAGGCTTTGGGCAAGAG 391
```

Db 31273 TGTGCTTCTGATATGACAGTAATATTATTCCTTCAGGCTCTAGTGGGTTGAAG 31214  
QY 392 CAGAG 396  
Db 31213 ACTAG 31209

RESULT 8  
US-10-719-993-6892/c  
; Sequence 6892, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6892  
; LENGTH: 51806  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-719-993-6892

Query Match 6.9%; Score 34.8; DB 18; Length 51806;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 316 CTTCAAGACGTAGTAGTACTCCACTAACTGAGTCCCATCTGAGAGTCACAGGAG 375  
Db 35334 CGTCAAGACGACGTATCTCAGTCAAGACGATCAGAACTGAGAGTAGTGGCAGCTG 35275

QY 376 GCTTTGGGGCAGAGCAGAGGAGGACCTGTTTGGTTGTTGAAGTTTGACTTAAC 435  
Db 35274 GAATTGTAGAAATATACACAGCAAGAAATGTCTATGAAGGACCCCTGATACTGTGAC 35215

QY 436 AAATTTGAGACATATGATGACATTGTGTGACACTTAACAACACTGACTCAGT 489  
Db 35214 TAAATTCTAAGCTGTACATACATACATACAGGGCTCTCAAAATTAACATAAG 35161

RESULT 9  
US-10-741-600-776/c  
; Sequence 776, Application US/10741600  
; Publication No. US20050026169A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001499  
; CURRENT APPLICATION NUMBER: US/10/741,600  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 73997  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 776  
; LENGTH: 43352  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(4352)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-741-600-776

Query Match 6.9%; Score 34.6; DB 19; Length 43352;  
Best Local Similarity 53.3%; Pred. No. 7.8;  
Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 352 TTCCCATCTGAGATCCACAGAGGCTTTGGGCAAGAGCAGAGGAGGAGGACTGTTTG 411

Db 4140 TTATCATCATATTCCTCAAGCAGGGTTACATGATAGAAAGAAAGAGGCTTCACTGC 4081  
QY 412 TGTGTGAAGTTTGTGCTCTTACAAATTGGAAGACATAGATGACATGTGTGAGACTAA 471  
Db 4080 TTTTTCATCTGTGAGACTTAAATGACCTGAAATTAATACATGAGAGTGTGTTTCTCA 4021

QY 472 CAACACCTAGACTCAT 488  
Db 4020 AAATGAATTACCAAAAT 4004

RESULT 10  
US-10-424-599-80920  
; Sequence 80920, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 80920  
; LENGTH: 634  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_44087C.1  
US-10-424-599-80920

Query Match 6.9%; Score 34.4; DB 17; Length 634;  
Best Local Similarity 60.9%; Pred. No. 3.4;  
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 142 CCTGGGTGACCTTGGTGGATATCTTGCTAACCTTAGATATTTCTTTGAGATATC 201  
Db 341 CTTGGGGAAATTTTGGGAGGAGATTTTGGTAAACCTTCTATATGTTTTTTTGAAGACT 400

QY 202 TTTAGAACATCCAGAAATATGGGCTTGATTTCT 233  
Db 401 ATTATGACATCGCCGACCAAGTTGGGATTTT 432

RESULT 11  
US-10-437-963-40327  
; Sequence 40327, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 40327  
; LENGTH: 4881  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_43780C.1  
US-10-437-963-40327



DB 881 TAACATACCTTAATTTAATCTTAATGACATTTTCCCTTACTTAAGCTTAGCA 822  
QY 275 GATCGTCAGAGAAACAGAGATCTCTCGTGTGATCTTCTTCA 321  
DB 821 ACTCTGAAAAAAACAAATGATGCTTTATTATTAGTAGCTTATTAGAA 775

## RESULT 15

US-10-087-192-331/c  
; Sequence 331, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 331  
; LENGTH: 295096  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(295096)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-331

## Query Match

6.7%; Score 33.6; DB 13; Length 295096;  
Best Local Similarity 48.0%; Pred. No. 1.4e+02;  
Matches 96; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 171 GGCTACCTTAGATATTTCTTTGGAGATCTTTAGAACATCCAGAAATAGGGCTTGAT 230  
DB 246768 GACTTTCTTGATGATCATTTAGACTAGATCTACCAACATTTGAAATGATGCCAGAT 246709  
QY 231 TCTCATCTCGGACCAATATTAAGTCACTAGATCCGAGATCGTGCAAGAAAC 290  
DB 246708 TTTGAAAAATGGAACTAAAAAGGCCATTTTGAAGCAATTAGCTCAAGTAAAAATA 246649  
QY 231 AAGATCTCTCTCGTGTGATCTTCTCAAGCAGTGAAGTAGTACTCCACTAACTGA 350  
DB 246648 AAAAATATGATGATCCAGAAAACCTCTGTATTGGCTAAAGTGTCACTCAACAGCTCTGA 246589  
QY 351 GTTCCATCTGAGAGTCCAC 370  
DB 246588 GTTGAAAATCTCAACTCCTC 246569

Search completed: February 10, 2005, 18:17:03  
Job time : 371.063 secs